THE AMERICAN MATHEMATICAL MONTHLY

Volume 106, Number 3 March 1999

Junpei Sekino — n-Ellipses and the Minimum Distance Sum Problem — 193

Arcadii Z. Grinshpan — The Bieberbach Conjecture and Milin's Functionals — 203

Dimitris A. Sardelis, Theodoros M. Valahas — Decision Making: A Golden Rule — 215

Vilmos Totik — A Tale of Two Integrals — 227

Jaume Paradís, Pelegrí Viader, Lluís Bibiloni — A Mathematical Excursion: From the Three-Door Problem to a Cantor-Type Set — 241

NOTES

Poh Wah Awyong — An Inequality Relating the Circumradius and Diameter of Two-Dimensional Lattice-Point-Free Convex Bodies — 252

Sherman K. Stein — Cutting a Polyomino into Triangles of Equal Areas — 255

William Staton — A Short Proof of Turán's Theorem — 257

Sung Soo Kim — A Characterization of the Set of Points of Continuity of a Real Function — 258

THE EVOLUTION OF...

I. G. Bashmakova, G. S. Smirnova — The Literal Calculus of Viète and Descartes — 260

PROBLEMS AND SOLUTIONS — 264

REVIEWS

Joseph H. Silverman — *An Accompaniment to Higher Mathematics.* By George R. Exner — 272

Journey Into Mathematics: An Introduction to Proofs. By Joseph Rotman — 272

Mathematical Thinking: Problem Solving and Proofs. By John P. D'Angelo and Douglas B. West — 272

John Meier, Trisha Thorme — *The Social Life of Numbers: A Quechua Ontology of Numbers and Philosophy of Arithmetic.* By Gary Urton — 275

TELEGRAPHIC REVIEWS — 280

EDITOR'S ENDNOTES — 284

AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1515 Mineral Square, Room 142
University of Utah
Salt Lake City, UT 84112

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper.

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues and descriptive summaries of forthcoming articles:

http://www.maa.org/

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

Reprint permission:
DONALD ALBERS, Director of Publication

Advertising Correspondence:
Mr. JOE WATSON, Advertising Manager

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center
maahq@maa.org

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Second class postage paid at Washington, DC, and additional mailing offices. ***Postmaster***: Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

n-Ellipses and the Minimum Distance Sum Problem

Junpei Sekino

1. INTRODUCTION. When we hear *ellipse*, we might think of planetary orbits or rooms with magical acoustic properties. In all such examples of ellipses in nature, the *foci* play a distinguished role. Our goal is to consider a natural class of generalized ellipses given by an arbitrary number of foci. Let $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ be n distinct points in the plane, and let k be a positive constant. By the *n-ellipse with foci* $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ and the *distance sum k*, we mean the level set of the *distance sum function*

$$f(\mathbf{r}) = \sum_{i=1}^{n} |\mathbf{r} - \mathbf{c}_i| \tag{1}$$

at the level $f(\mathbf{r}) = k$. We show that if k is sufficiently large (as explained in Theorem 6), then the n-ellipse is a piecewise smooth Jordan curve whose interior is convex; it is nonsmooth only where it contains a focus. The n-ellipses have diverse shapes that include curves resembling contours of eggs, lemons, pears, and even human faces, symmetric or asymmetric. Circles are 1-ellipses and ordinary ellipses are 2-ellipses.

Surveying a family of n-ellipses given by a set of n foci on a computer screen reveals not only a variety of continuously changing contours but also the existence of a *center*, which is the *unique* point that minimizes the distance sum function; see Figure 1. It is interesting to explore the general behavior of the distance sum function but it is also important to examine its local properties near the center. The latter leads us to the critical points of the distance sum function and to

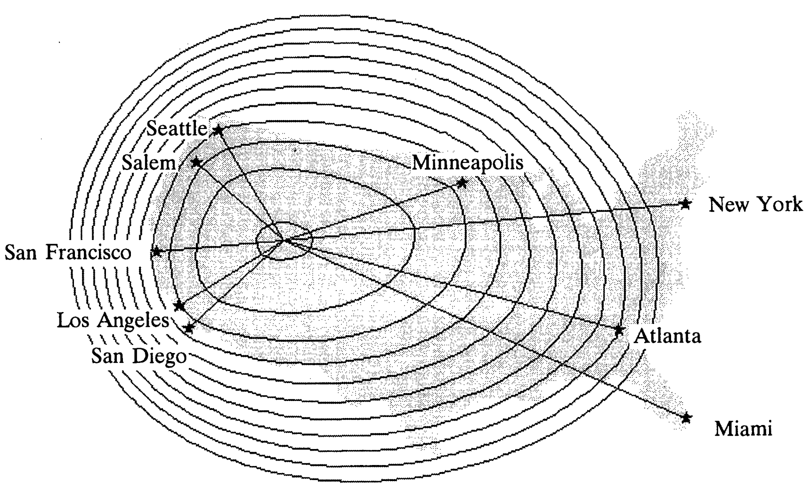


Figure 1. A family of 9-ellipses and a center

geometrically intriguing properties. We show that a critical point is nondegenerate if and only if the foci are noncollinear. This theorem, which places circles and ordinary ellipses in a minor degenerate league, shows that the geometrically and analytically appealing case arises when the foci do not lie in a straight line. The existence of a center under this condition follows quickly.

One outcome of our exploration is a mathematical model for certain optimization problems, and the final section of this paper lists a few examples that can be solved by a contour map and the properties of critical points. These problems may be given to students in a second year calculus or optimization class as a challenging project.

2. CONTOUR PLOTTING AND EXAMPLES. Sherlock Holmes said, "It is a capital crime to theorize before one has data." To gather data and develop our intuition, we used a contour plotter to draw a family of n-ellipses on the computer screen and display an approximate location of the center. We wrote a program to perform the task **[4]**; commercial contouring routines are available in computer algebra systems such as *Maple* and *Mathematica*.

Example 1. Figure 1 shows a family of 9-ellipses and its center, which is related to a project in the final section. It shows that an n-ellipse need not contain all foci (indicated by stars) in its interior when $n > 2$.

Example 2. As Figure 2 shows, a 3-ellipse generated by the vertices of an isosceles triangle can resemble the familiar section of an egg. The vertices of a rectangle generate 4-ellipses that resemble an ordinary ellipse, and so does a finite set of equally spaced collinear points.

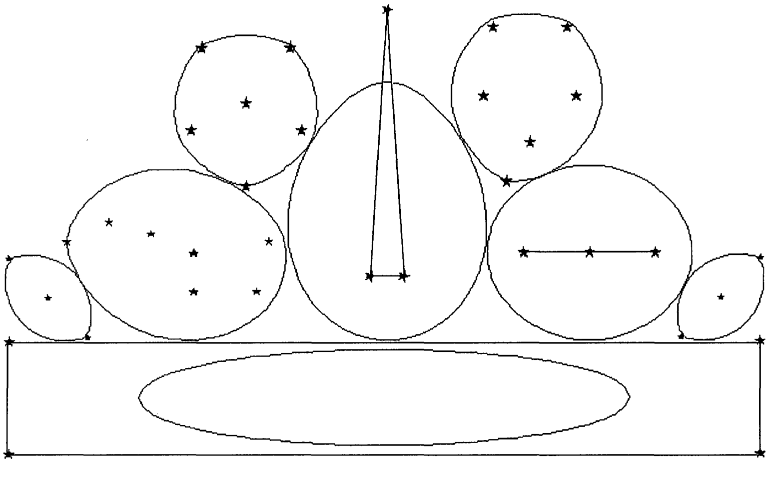


Figure 2. Assorted n-ellipses

3. PRELIMINARY RESULTS. Consider the simple distance function

$$g(\mathbf{r}) = |\mathbf{r} - \mathbf{c}| = \sqrt{(x-p)^2 + (y-q)^2} \tag{2}$$

centered at a point $\mathbf{c} = (p, q)$ where $\mathbf{r} = (x, y)$. The domain of definition of g is the entire plane, and g is a C^∞-function on the plane with the focus $\mathbf{c}$ removed. A

straightforward calculation gives $\nabla|\mathbf{r}-\mathbf{c}| = (\mathbf{r}-\mathbf{c})/|\mathbf{r}-\mathbf{c}|$ for $\mathbf{r} \neq \mathbf{c}$, so $\nabla g(\mathbf{r})$ is the (unit) direction vector from $\mathbf{c}$ to $\mathbf{r}$. We shall employ the convenient notation $\nabla|\mathbf{b}-\mathbf{a}|$ for the direction from a point $\mathbf{a}$ to another point $\mathbf{b}$. By the linearity of ∇, we have:

Theorem 1. *The distance sum function f in* (1) *is C^∞ on the plane with the foci removed, and*

$$\nabla f(\mathbf{r}) = \sum_{i=1}^{n} \nabla|\mathbf{r}-\mathbf{c}_i| = \sum_{i=1}^{n} \frac{\mathbf{r}-\mathbf{c}_i}{|\mathbf{r}-\mathbf{c}_i|}.$$

Thus, ∇f is a gradient field on the plane with holes at the foci; we draw the gradient $\nabla f(\mathbf{r})$ by placing its tail at $\mathbf{r}$. According to Theorem 1, each $\nabla f(\mathbf{r})$ is completely determined by the directions $\nabla|\mathbf{r}-\mathbf{c}_i|$, $i = 1, 2, \ldots, n$, which we call the *direction components* of $\nabla f(\mathbf{r})$. Figure 3 illustrates a 2-ellipse and a 3-ellipse with gradients and their direction components (bold arrows) and the tangent lines perpendicular to the respective gradients. The 2-ellipse shows that the line segments $\mathbf{rc}_1$ and $\mathbf{rc}_2$ make equal angles with the tangent line. This "ball bouncing property" of an ordinary ellipse from one focus to another off the wall is not available in n-ellipses if $n > 2$.

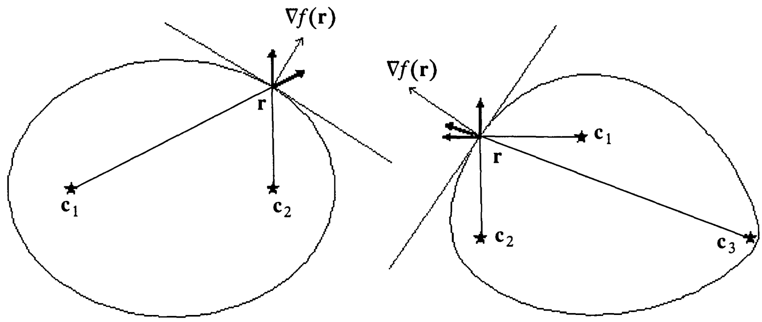


Figure 3. A 2-ellipse and a 3-ellipse

The next lemma implies that every vertical section of the distance sum function f has *at most* one "valley" corresponding to a local minimum and no "hill" that corresponds to a local maximum.

Lemma 1. *Let L be any line, and parametrize L by $\mathbf{r}(t) = \mathbf{d}t + \mathbf{b}$ where $\mathbf{d}$ is a unit vector. Then $f(\mathbf{r}(t))$ is continuous and the directional derivative of f along L is piecewise continuous and monotone increasing* (*i.e.*, *nondecreasing*).

Proof: First, consider the distance function g in (2). The directional derivative $\nabla g(\mathbf{r}(t)) \cdot \mathbf{d}$ is the scalar projection of the direction vector $\nabla|\mathbf{r}(t)-\mathbf{c}|$ onto $\mathbf{d}$ and therefore it is increasing if $\mathbf{c} \notin L$ (see Figure 4). Similarly, if $\mathbf{c} \in L$, $\nabla g(\mathbf{r}(t)) \cdot \mathbf{d}$ is a monotone increasing step function that is undefined when $\mathbf{r}(t) = \mathbf{c}$. Since $f(\mathbf{r}(t))$ is a finite sum of functions of the form $g(\mathbf{r}(t))$, $f(\mathbf{r}(t))$ is continuous and $\nabla f(\mathbf{r}(t)) \cdot \mathbf{d}$ is monotone increasing. The directional derivative does not exist when $\mathbf{r}(t)$ coincides with a focus of f. ■

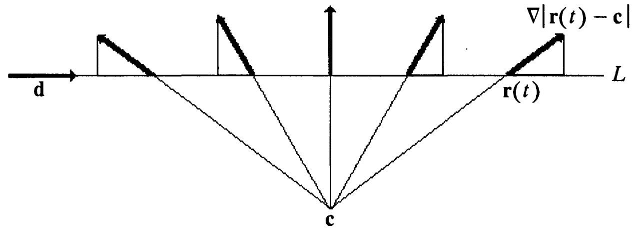


Figure 4. The directional derivative of g along L

Theorem 2. *The function f in* (1) *has a global minimum.*

Proof: Choose a compact disk D containing all the foci in its interior, and let C be the boundary of D. Then f attains a global minimum value M on D at some point $\mathbf{s} \in D$. Theorem 1 ensures that $\nabla f(\mathbf{r}) \neq \mathbf{0}$ for all $\mathbf{r} \in C$, and therefore, $\mathbf{s}$ belongs to the interior of D. Let L be any ray in the plane emanating from $\mathbf{s}$, and let $\mathbf{b}$ be the intersection of L and C. Parametrize L by $\mathbf{r}(t) = \mathbf{d}t + \mathbf{s}$, $t \geq 0$, where $\mathbf{d} = \nabla|\mathbf{b} - \mathbf{s}|$. Then Lemma 1 and $f(\mathbf{s}) = M$ imply $\nabla f(\mathbf{b}) \cdot \mathbf{d} \geq 0$. Appealing to Lemma 1 once again, therefore, we have $M = f(\mathbf{s}) \leq f(\mathbf{r}(t))$ for all $t \geq 0$. This proves that M is the global minimum value of f on L, and the theorem follows. Note that M need not be achieved at a unique point. ■

4. CRITICAL POINTS. By a *critical point* (abbreviated *CP*), we mean a point $\mathbf{r}$ where $\nabla f(\mathbf{r}) = \mathbf{0}$. This includes the assumption that $\mathbf{r}$ is not a focus. We say that a CP is *with n* foci if f has exactly n foci.

Theorem 1 ensures that $\mathbf{r}$ is a CP if and only if

$$\sum_{i=1}^{n} \nabla|\mathbf{r} - \mathbf{c}_i| = \sum_{i=1}^{n} \frac{\mathbf{r} - \mathbf{c}_i}{|\mathbf{r} - \mathbf{c}_i|} = \mathbf{0}. \tag{3}$$

While this formula discourages algebraic approaches to a solution, it gives a striking geometric property of a CP: *The directions from the foci add up to zero precisely at a CP, whence a CP depends only upon the directions from the foci but not the distances.* Figure 5 stresses this point and shows distinct sets of foci (one set for

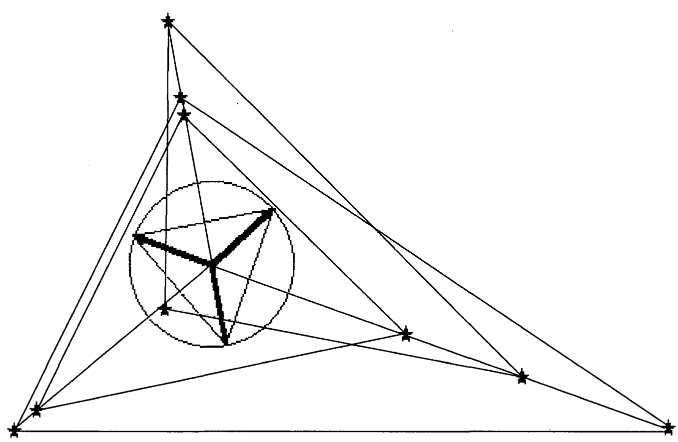


Figure 5. Relationship between the foci and a CP

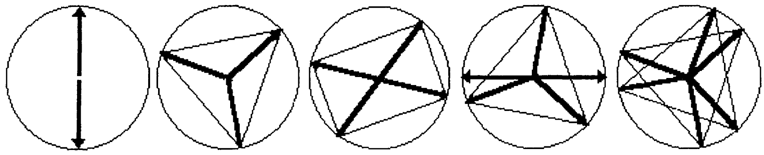


Figure 6. Examples of CP-patterns with 2, 3, 4, 5, and 6 foci

each triangle) that have the same CP (the center of the unit circle). The circular part of Figure 5, which we call a *CP-pattern*, consists of a CP and the direction components (bold arrows) of $\nabla f(\mathbf{r})$ whose sum vanishes at the CP. Note that the *negative* of each direction component $\nabla |\mathbf{r} - \mathbf{c}_i|$ points toward the focus $\mathbf{c}_i$. Figure 6 illustrates a few more CP-patterns. The first CP-pattern of Figure 6 describes the only way ∇f can vanish when there are exactly two foci, and therefore $\mathbf{r}$ is a CP with two foci if and only if $\mathbf{r}$ is strictly between the foci; the second CP-pattern shows that $\mathbf{r}$ is a CP with three foci if and only if the angle between any pair of the direction components is 120°; the third pattern shows that $\mathbf{r}$ is a CP with four noncollinear foci if and only if $\mathbf{r}$ is the center of the quadrilateral formed by the foci. The idea of locating CPs by means of the CP-patterns becomes elusive, however, if $n \geq 5$. The fourth pattern shows a CP with five foci, and the rotations of the horizontal vectors bound together yield infinitely many distinct CP-patterns with five foci. In addition, a regular pentagon gives yet another CP-pattern with five foci, and there are others that match none of the above. The fifth pattern is just one of the infinitely many CP-patterns with six foci.

The implication from the CP-pattern with two foci just observed can be extended easily to the following:

Theorem 3. *Suppose the foci of an n-ellipse are collinear. If n is even then a point* $\mathbf{r}$ *is a CP if and only if* $\mathbf{r}$ *lies strictly between the middle two foci. If n is odd, no CPs exist.*

According to the theorem, the distance sum function f can have *no* CPs or *infinitely many* CPs, and this raises the question: If the foci are noncollinear, how many CPs can f have? The second and third patterns in Figure 6 indicate that f can have at most one CP if the number of foci is 3 or 4, but it is not clear what happens for more foci. To settle this question, therefore, we take an analytic route. We say that a CP $\mathbf{r}$ is *degenerate* if

$$Hf(\mathbf{r}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial y\, \partial x} \\ \dfrac{\partial^2 f}{\partial x\, \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix},$$

the *Hessian matrix* of f at $\mathbf{r}$, is singular.

Theorem 4. *A CP of an n-ellipse is nondegenerate if and only if the foci are noncollinear. Furthermore, every nondegenerate CP is a local minimum of the distance sum function.*

Proof: Let $\mathbf{r} = (x, y)$ and $\mathbf{c}_i = (p_i, q_i)$. To avoid cluttered formulas, let

$$X_i = x - p_i, \quad Y_i = y - q_i, \quad \text{and} \quad g_i = g_i(\mathbf{r}) = |\mathbf{r} - \mathbf{c}_i| = \sqrt{X_i^2 + Y_i^2}\,.$$

Then

$$\frac{\partial g_i}{\partial x} = \frac{X_i}{g_i}, \qquad \frac{\partial g_i}{\partial y} = \frac{Y_i}{g_i},$$

and

$$\nabla f(\mathbf{r}) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = \left(\sum \frac{\partial g_i}{\partial x}, \sum \frac{\partial g_i}{\partial y}\right) = \left(\sum \frac{X_i}{g_i}, \sum \frac{Y_i}{g_i}\right),$$

where each summation is taken over $i = 1, 2, \ldots, n$. Hence,

$$\frac{\partial^2 f}{\partial x^2} = \sum \frac{\partial}{\partial x}\frac{X_i}{g_i} = \sum \frac{\frac{\partial X_i}{\partial x} g_i - X_i \frac{\partial g_i}{\partial x}}{g_i^2} = \sum \frac{g_i - \frac{X_i^2}{g_i}}{g_i^2} = \sum \frac{g_i^2 - X_i^2}{g_i^3} = \sum \frac{Y_i^2}{g_i^3},$$

and similarly,

$$\frac{\partial^2 f}{\partial y\, \partial x} = \frac{\partial^2 f}{\partial x\, \partial y} = -\sum \frac{X_i Y_i}{g_i^3}, \quad \text{and} \quad \frac{\partial^2 f}{\partial y^2} = \sum \frac{X_i^2}{g_i^3}.$$

Setting $A_i = X_i / g_i^{3/2}$ and $B_i = Y_i / g_i^{3/2}$, the Hessian matrix of f can be written as

$$Hf(\mathbf{r}) = \begin{bmatrix} \sum B_i^2 & -\sum A_i B_i \\ -\sum A_i B_i & \sum A_i^2 \end{bmatrix}.$$

Therefore,

$$\det Hf(\mathbf{r}) = \left(\sum A_i^2\right)\left(\sum B_i^2\right) - \left(\sum A_i B_i\right)^2 = |A|^2 |B|^2 - (A \cdot B)^2$$

where $A = (A_1, A_2, \ldots, A_n)$ and $B = (B_1, B_2, \ldots, B_n)$. It follows from the Cauchy-Schwarz inequality that $\det Hf(\mathbf{r}) \geq 0$, and $\det Hf(\mathbf{r}) = 0$ if and only if A and B are linearly dependent. The rest follows from the second derivative test. ■

According to Theorem 4, the CPs with 3, 4, 5, and 6 foci in Figure 6 are all nondegenerate. Since nondegenerate CPs are isolated [**3**, p. 8], Theorem 4 and Lemma 1 imply

Corollary 1. *If the foci of an n-ellipse are noncollinear, the distance sum function f has at most one CP, i.e., a CP is unique if it exists.*

We say that a point $\mathbf{s}$ is the *center* of the distance sum function f if $\mathbf{s}$ is the *unique* point at which f attains a global minimum. We now show that f has a center unless the foci are collinear and the number of foci is even.

Theorem 5. *Let an n-ellipse be given. (A) Suppose the foci are noncollinear. If a CP exists then it is the center; otherwise, one of the foci coincides with the center. (B) Suppose the foci are collinear. If n is even, then f has no center, and instead, f attains a global minimum at $\mathbf{r}$ if and only if $\mathbf{r}$ lies in the closed line segment joining the middle two foci; if n is odd, then the middle focus is the center of f.*

Proof: Let S be the set consisting of all CPs and foci and let M be the global minimum value of f. Since every local minimum of f is either a CP or a point at which f is not differentiable, it follows that f attains the value M at some point $\mathbf{s} \in S$ but never at a point in the complement of S.

(A) S is finite by Corollary 1, and Lemma 1 guarantees uniqueness of the point $\mathbf{s}$ where $f(\mathbf{s}) = M$, i.e., $\mathbf{s}$ is the center of f. Now, if no CPs exist then $\mathbf{s}$ must be a focus; if there is a CP at $\mathbf{r}$ then Theorem 4 ensures that $f(\mathbf{r})$ is a local minimum; appealing to Lemma 1 again, we conclude that $\mathbf{r} = \mathbf{s}$.

(B) Suppose n is even, and let L be the line through the foci. Then $S \subset L$ by Theorem 3, and consequently the global minimum value of f over L coincides with M. Let $\mathbf{d}$ be a direction of L and suppose without loss of generality that the foci are lined up on L in the direction $\mathbf{d}$ as

$$\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_j, \mathbf{c}_{j+1}, \dots, \mathbf{c}_{2m} \quad \text{so that} \quad \mathbf{d} = \nabla|\mathbf{c}_2 - \mathbf{c}_1|. \tag{4}$$

Parametrize L by $\mathbf{r}(t) = \mathbf{d}t + \mathbf{b}$, and given any index j, choose $t = t_j$ in such a way that $\mathbf{r}_j = \mathbf{r}(t_j)$ is strictly between $\mathbf{c}_j$ and $\mathbf{c}_{j+1}$ on L. Then $\mathbf{r}_j$ separates L into two rays R_1 and R_2 such that $\mathbf{c}_j \in R_1$, and therefore R_1 and R_2 contain exactly j and $2m - j$ foci, respectively. Theorem 1 and (4) ensure that

$$\begin{aligned} \nabla f(\mathbf{r}_j) &= \sum_{i=1}^{j} \nabla|\mathbf{r}_j - \mathbf{c}_i| + \sum_{i=j+1}^{2m} \nabla|\mathbf{r}_j - \mathbf{c}_i| \\ &= \sum_{i=1}^{j} \nabla|\mathbf{c}_2 - \mathbf{c}_1| + \sum_{i=j+1}^{2m} \nabla|\mathbf{c}_1 - \mathbf{c}_2| = \sum_{i=1}^{j} \mathbf{d} - \sum_{i=j+1}^{2m} \mathbf{d} = 2(j-m)\mathbf{d}, \end{aligned}$$

and $\nabla f(\mathbf{r}_j) \cdot \mathbf{d} = 2(j-m)|\mathbf{d}|^2 = 2(j-m)$. Consequently, $d(f(\mathbf{r}(t)))/dt = \nabla f(\mathbf{r}) \cdot \mathbf{d} = 0$ if and only if $\mathbf{r}(t)$ is strictly between $\mathbf{c}_m$ and $\mathbf{c}_{m+1}$. Since $f(\mathbf{r}(t))$ is continuous, we now conclude that $f(\mathbf{r}(t)) = M$ if and only if $\mathbf{r}(t)$ lies in the closed line segment joining the middle two foci. A similar argument proves the second part of (B). ■

Although Theorem 5(A) guarantees the existence of a (unique) center under any noncollinear arrangement of the foci, it is still incomplete in the sense that it tells neither when a CP exists nor which focus is the center if no CPs exist. Theorem 5(A) can be strengthened as follows:

Suppose the foci are noncollinear and therefore a CP is unique whenever it exists (Corollary 1). From the second and third CP-patterns in Figure 6, we have:

(C) There is a unique CP with three foci if and only if the foci form a triangle whose interior angle never exceeds or equals 120°.

(D) There is a unique CP with four foci if and only if the foci form a convex quadrilateral whose interior angle never equals 180° .

If a CP with three foci fails to exist, which focus is the center? The answer is the focus that corresponds to the greatest interior angle of the triangle, which can be verified easily by checking the directional derivative of f on the line segments joining it and the other foci. The answer is the same for four foci. For more foci, CPs abound:

(E) There is a unique CP with n foci if the foci form a convex n-gon whose interior angle never equals 180°.

The sufficient condition in (E) is not necessary: We can easily build a counterexample using the CP-pattern with five foci in Figure 6. To justify (E), suppose $n = 5$

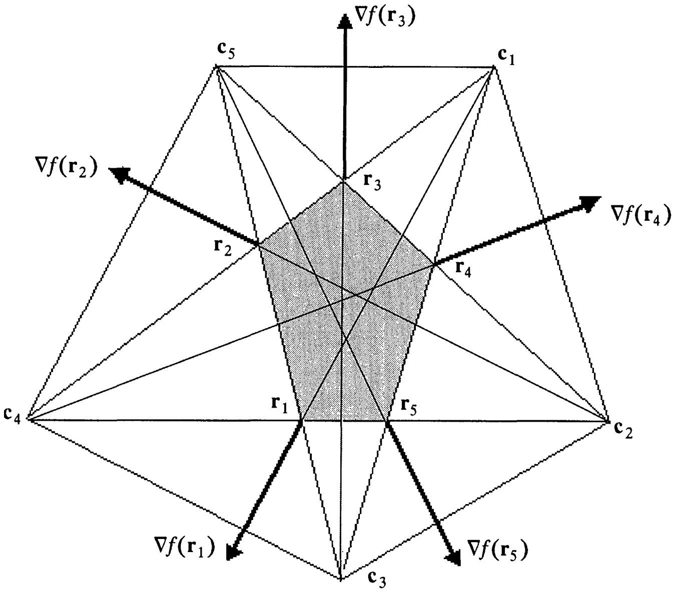


Figure 7. Is the center inside the little pentagon?

and consider the convex pentagon in Figure 7. Each $\mathbf{r}_i$ is a vertex of the smaller pentagon constructed by the "star-forming" diagonals. Let P be the smaller pentagon and let L be the closed line segment joining adjacent vertices of P, say $\mathbf{r}_1$ and $\mathbf{r}_2$. Each gradient $\nabla f(\mathbf{r}_i)$, $i = 1, 2$, is given by offsetting four of its direction components as shown in the picture, and as a result,

$$-\nabla f(\mathbf{r}_i) \text{ points into the interior of } P, \tag{5}$$

i.e., $-\nabla f(\mathbf{r}_i)$ has a positive projection on the inward unit normal to L at $\mathbf{r}_i$.

Parametrize L by $\mathbf{r}(t) = \mathbf{d}t + \mathbf{r}_1$, where $\mathbf{d} = \nabla |\mathbf{r}_2 - \mathbf{r}_1|$. Since L does not contain any focus, the directional derivative $\nabla f(\mathbf{r}(t)) \cdot \mathbf{d}$ is continuous, and Lemma 1 ensures that its negative

$$-\nabla f(\mathbf{r}(t)) \cdot \mathbf{d} = \cos\theta(\mathbf{r}(t))$$

is monotone decreasing, where $\theta(\mathbf{r}(t)) = \cos^{-1}(-\nabla f(\mathbf{r}(t)) \cdot \mathbf{d})$ is the angle between $-\nabla f(\mathbf{r}(t))$ and $\mathbf{d}$. Consequently, $\theta(\mathbf{r}(t))$ is monotone increasing from $\theta(\mathbf{r}_1)$ to $\theta(\mathbf{r}_2)$ where $0° < \theta(\mathbf{r}_1) < \theta(\mathbf{r}_2) < 180°$, and the vector field $-\nabla f$ points into the interior of P along the line segment L (except possibly at the CP if indeed L contains it). This implies that P contains the center, which in turn coincides with the CP of f (Theorem 5(A)).

If $n > 5$, $\nabla f(\mathbf{r}_i)$ is the sum of $n - 4$ direction components but the preceding argument carries over as the property (5) can be observed easily under the general circumstance.

Our final theorem concerns the general shape of n-ellipses.

Theorem 6. *Let M be the global minimum value of f. Every n-ellipse with the distance sum greater than M is a piecewise smooth Jordan curve and its interior is a nonempty convex set.*

Proof: Given $k > M$, let E be the n-ellipse with $f(\mathbf{r}) = k$. Let $S = f^{-1}[M, k]$ and let int(S) denote the interior of S. Then int$(S) = f^{-1}[M, k)$, which is nonempty.

To show the convexity of int(S), let $\mathbf{r}_1, \mathbf{r}_2 \in \text{int}(S)$, and parametrize the line L through $\mathbf{r}_1, \mathbf{r}_2$ by $\mathbf{r}(t) = \mathbf{d}t + \mathbf{r}_1$. Then Lemma 1 ensures that $f(\mathbf{r}) \leq \max\{f(\mathbf{r}_1), f(\mathbf{r}_2)\} < k$, whenever $\mathbf{r}$ is in L between $\mathbf{r}_1$ and $\mathbf{r}_2$. This means that the portion of L between $\mathbf{r}_1$ and $\mathbf{r}_2$ is in int(S), so int(S) is convex. Next, we show that S is bounded, and therefore it is compact. Theorem 5 guarantees that there is a point $\mathbf{s}$ such that $f(\mathbf{s}) = M$. If $\mathbf{r} \in S$ then $|\mathbf{r} - \mathbf{s}| \leq |\mathbf{r} - \mathbf{c}_i| + |\mathbf{s} - \mathbf{c}_i|$ for each focus $\mathbf{c}_i$, so

$$n|\mathbf{r} - \mathbf{s}| \leq \sum_{i=1}^{n} |\mathbf{r} - \mathbf{c}_i| + \sum_{i=1}^{n} |\mathbf{s} - \mathbf{c}_i| = f(\mathbf{r}) + f(\mathbf{s}) \leq k + M.$$

This proves that S is bounded. Now Theorem 5 also guarantees that all CPs are in int(S) whenever they exist. Therefore the Implicit Function Theorem ensures that E is a piecewise smooth curve: If no focus lies in E, then E is smooth; otherwise, cusps may occur at the foci that connect the smooth segments given by the Implicit Function Theorem. The n-ellipse E bounds a convex and compact region S, whence it must be a Jordan curve whose interior coincides with int(S). ∎

5. STUDENT PROJECTS. The following problems are open-ended and solutions need not be unique. Your report must describe the logic behind your solutions and must explain how you used computers or calculators. Use the map provided (Figure 1 without the curves and lines).

A woman owns a trendy specialty-food store in each of the nine cities shown in the map.

(a) She plans to open another store in the United States. The new site is to be her headquarters and she wants the location to be the most convenient for frequent trips she makes by her private airplane to the nine cities. If she visits each city equally frequently directly from the headquarters, where should she locate the new store?

In the following, suppose you have solved the nine city problem and that the owner of the chain has built her headquarters.

(b) She wants to open two more stores including one in San Antonio, but she does not want to move her headquarters. What other city should she choose? What if "two" were replaced by "three"? "four"? "five"?

(c) If she wants to relocate the Atlanta store without moving the headquarters, where should she move it?

(d) If changing American eating habits force her to close two of the nine stores, but she does not want to relocate her headquarters, which two stores should she close?

(Possible Solutions) The contour plot of Figure 1, which is given by assigning a coordinate system on the map, points to a mountainous area near the Utah-Wyoming border: (a) Choose Salt Lake City, the largest city in the area. From the geometric interpretation of (3), we also have: (b) Boise, Idaho, which "counterbalances" San Antonio with respect to the center; (c) Choose a city on the line segment between Salt Lake City and Atlanta; Yes, moving the store from Atlanta to, say Denver, changes the contours but the center stays miraculously in the same area. (d) San Francisco and New York.

Inquisitive students may wish to extend the idea of n-ellipse to that of *weighted n-ellipse* given by the new function

$$f(\mathbf{r}) = \sum_{i=1}^{n} w_i|\mathbf{r} - \mathbf{c}_i|$$

with *weights* $w_i > 0$. The critical points $\mathbf{r}$ of f satisfy the equation $\sum_{i=1}^{n} w_i \nabla |\mathbf{r} - \mathbf{c}_i| = 0$. This causes the corresponding CP-patterns to become more complex, but as a trade-off, the weights provide us with optimization problems that are more realistic. The following example shows the effect of weights on the center:

(e) Because of the varying business size, the owner of the chain in (a) above put the following weights on the nine cities: Seattle = 2.5, Salem = 1, SF = 2.4, LA = 2.2, SD = 2.4, Miami = 2.2, Atlanta = 2, NY = 6, and Minneapolis = 2. This means, for example, that she visits NY six times as frequently as Salem, Oregon. Where should she locate her headquarters? Figure 8 shows that the center shifts

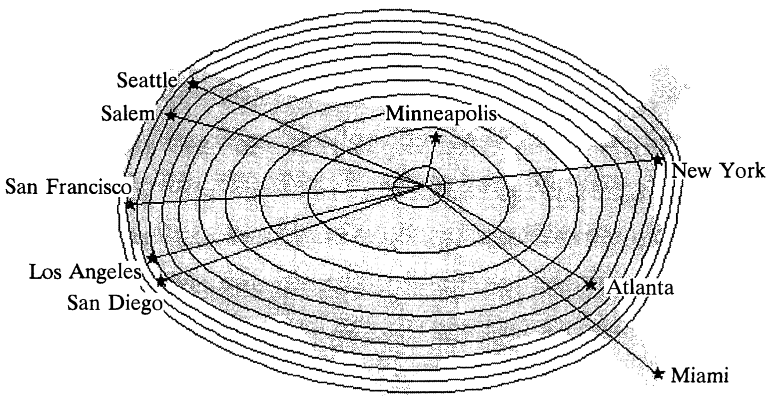


Figure 8. A family of weighted 9-ellipses

eastward to a point near Iowa City mainly due to the heavy weight placed on New York.

REFERENCES

1. W. Kaplan, *Advanced Calculus*, 4th ed., Addison Wesley, Reading, 1991.
2. J. E. Marsden and A. J. Tromba, *Vector Calculus*, 4th ed., Freeman, San Francisco, 1996.
3. J. W. Milnor, *Morse Theory*, Princeton University Press, Princeton, 1969.
4. J. Sekino, N. Liepins, and V. Morachevsky, Contour Plot for Smooth Surfaces, Proceedings of St. Petersburg State University (Astronomy/Mathematics/Mechanics Branch) **4** (22) (1994), 117–118.

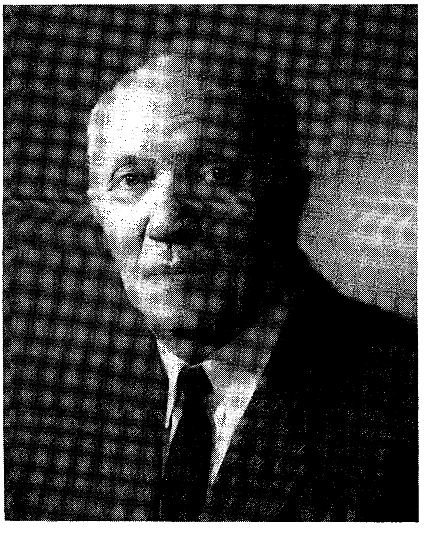
JUNPEI SEKINO finished an undergraduate program in chemistry at Nihon University, Tokyo, and came to Oregon State University, where he was influenced by teachers/mentors Gordon W. Gilkey, Ze'ev B. Orzech, and J. Wolfgang Smith. After receiving a second bachelor's degree in art, he earned his Ph.D. in algebraic topology under Smith. He is professor at Willamette University, where he was once the Mortarboard Professor of the Semester. His hobbies include wild mushroom hunting, printmaking, and fractal plotting. His prizewinning web site "Sekino's Fractal Gallery" can be accessed via the home page of the MAA's Pacific Northwest Section (http://www.math.ubc.ca/ ~ cayf/Pacnortsect.html).
Willamette University, Salem, OR 97301
sekino@willamette.edu

The Bieberbach Conjecture and Milin's Functionals

Arcadii Z. Grinshpan

Dedicated to the memory of Isaak Moiseevich Milin (1919–1992)

The now-settled Bieberbach conjecture (1916) for the Taylor coefficients of univalent (1–1 analytic) functions is one of the most famous and inspirational problems of mathematics. The following events are milestones in the long history of this problem.



(i) Bieberbach's conjecture motivated the development of the Loewner parametric method (1923), which became a very useful tool in the theory of univalent functions.

(ii) In 1971, I. M. Milin constructed a sequence of functionals associated with his exponentiation approach and the Bieberbach conjecture. He conjectured that these functionals were nonpositive, and gave an elementary argument showing that his conjecture implies Bieberbach's.

(iii) In 1984, L. de Branges proved that Milin's functionals were nonpositive. He used Loewner's method together with results and ideas from several fields of mathematics. This achievement allowed de Branges to confirm the Bieberbach conjecture as a theorem.

de Branges' original argument was complicated. During 1984–97, analysts from several countries simplified every component of his approach. Using the ideas of various authors, we prove the Bieberbach and Milin conjectures in a way accessible to many readers. A helpful basic reference is [**4**].

1. THE BIEBERBACH CONJECTURE: SIMPLE TO STATE BUT HARD TO PROVE.

According to the Riemann mapping theorem (1851) [**4**, Section 1.5], every simply connected domain D that is a proper subset of the complex plane can be mapped analytically and 1–1 onto the unit disk $E = \{z : |z| < 1\}$. Furthermore, there is a unique such mapping that takes a given point in D into the origin and has a positive derivative there. This fundamental result allows complex analysts to formulate many extremal problems in the plane in terms of normalized univalent (*schlicht*) functions in E. Therefore the analytic and geometric properties of such functions are of interest. The behavior of their Taylor coefficients has been the subject of intensive research for most of the Twentieth Century.

Let S (for *schlicht*) be the class of functions $f(z)$ that are analytic and 1–1 in E, and normalized by the conditions $f(0) = 0$, $f'(0) = 1$. An important example is the Koebe function $K(z) = z/(1-z)^2 = \sum_{n=1}^{\infty} nz^n$, which maps E onto the plane slit along $(-\infty, -1/4]$. This function is extremal for many classical functionals on S. For example, for all $f \in S$ and all $z \in E$ we have

$$-K(-|z|) \leq |f(z)| \leq K(|z|), \tag{1}$$

which is known as the *growth theorem* [**4**, Section 2.3]. The Koebe function is of particular interest when dealing with coefficient estimates.

Let $\{f\}_n$ denote the coefficient of z^n in the Taylor series expansion about $z = 0$ of a function $f(z)$. Note that $\{K\}_n = n$, $n = 1, 2, \dots$.

The Bieberbach conjecture (1916) [**2**] *asserts that*

$$|\{f\}_n| \leq n, \quad n = 2, 3, \dots \tag{2}$$

for each $f \in S$, and that equality holds for any given n only for the Koebe function $K(z)$ and its rotations $\bar{\lambda}K(\lambda z)$, $|\lambda| = 1$.

For nearly 70 years, many mathematicians, including the very best analysts, tried to prove or disprove this conjecture [**1**], [**4**], [**10**]. L. Bieberbach himself proved it for $n = 2$ [**2**]. Here is a short version of his proof based on the fact that area is always nonnegative.

Theorem A. (L. Bieberbach [**2**]). *If $f \in S$, then $|\{f\}_2| \leq 2$, with equality if and only if f is a rotation of the Koebe function.*

Proof: The function $F(z) = [f(z^2)]^{-1/2} = z^{-1} - \frac{1}{2}\{f\}_2 z + \cdots$ is univalent in $E \setminus \{0\}$ [**4**, Section 2.1]. For $r \in (0,1)$, let C_r be the image under F of the circle $|z| = r$. Clearly, C_r is a simple closed curve. Switching to polar coordinates, we write $F(re^{i\alpha}) = Re^{i\Psi}$, $0 \leq \alpha < 2\pi$. Since the area enclosed by C_r is positive, we have

$$\frac{1}{2}\int_{C_r} R^2 \, d\Psi > 0,$$

where the integration is performed along C_r in the counterclockwise sense. Using $R\Psi_\alpha = rR_r$, one of the Cauchy-Riemann equations in polar coordinates, we find

$d\Psi = (r/R)R_r\, d\alpha$. It follows that

$$-\frac{d}{dr}\int_0^{2\pi}\left|F(re^{i\alpha})\right|^2 d\alpha = 4\pi\left(r^{-3} - \left|\frac{1}{2}\{f\}_2\right|^2 r - \cdots\right) > 0.$$

As $r \to 1$ we deduce that $|\{f\}_2| \le 2$. Equality is possible only if $F(z) = z^{-1} - \lambda z$, where $|\lambda| = |\frac{1}{2}\{f\}_2| = 1$, and thus $f(z) = \bar{\lambda}K(\lambda z)$. ∎

The present proof of the Bieberbach conjecture for $n > 2$ is based on four theorems (A, B, C, D) and four lemmas (1, 2, 3, 4). The desired result is an immediate consequence of Milin's Theorem C and de Branges' Theorem D. Theorem C states that if certain functionals on S are nonpositive (the Milin conjecture) then the Bieberbach conjecture is true. Theorem D asserts the truth of the Milin conjecture.

We follow Milin's argument to prove Theorem C. A particular case of his monotonicity lemma (Lemma 2) implies the so-called Lebedev-Milin exponential inequalities for formal power series. We obtain (2) by applying the Cauchy-Schwarz inequality and the Lebedev-Milin inequalities to the coefficients of a function in S. Theorem A is used to settle the case of equality.

The proof of Theorem D based on de Branges' idea requires some familiarity with Loewner's approach (Section 2). It is sufficient to prove the nonpositivity of Milin's functionals only for a dense subclass of S (Lemma 1) associated with Loewner's differential equation (3) (Theorem B). We use a simplified coefficient form of de Branges' construction to produce for each f in this subclass, and for each Milin functional I, a differentiable function $\varphi(t)$, $0 \le t \le T$, so that $\varphi(0) = I(f)$ and $\varphi(T) = 0$. We combine Loewner's equation, certain polynomial inequalities (Corollary of Lemma 3), and a simple lemma (Lemma 4) to show that $\varphi' \ge 0$; it follows that $I(f) \le 0$.

All theorems and lemmas are arranged in historical order, which also turns out to be the logical one. Thus, a certain amount of energy and patience allows the reader to walk in the footprints of C. Loewner, I.M. Milin, L. de Branges, and many other mathematicians in proving the famous conjecture.

2. THE LOEWNER METHOD: A POWERFUL BUT SELECTIVE TOOL. The parametric method introduced by C. Loewner in 1923 [**8**] and later developed by other authors (see [**4**, Chapter 3]) permits one to solve many extremal problems on the class S via reduction to a dense subclass associated with a partial differential equation.

Although Loewner's method has been very productive, it is far from being universal. Loewner used it to prove (2) for $n = 3$ (and $n = 2$) in his original paper [**8**]. However a proof for $n = 4$ based solely on Loewner's method was given (by Z. Nehari) only 50 years later, when the cases $n = 4, 5, 6$ had been settled by other means (see [**4**; Sections 3.5, 4.6 and Notes, pp. 69, 139]). Despite substantial efforts, no one was able to use Loewner's method in a direct proof of any case $n > 4$. In fact, prior to 1984 no method yielded the conjectured estimate for coefficients beyond the first six.

Fortunately, Loewner's representation theorem for single-slit mappings (functions that map E onto the plane slit along a Jordan arc) can be applied to Milin's functionals. It is a consequence of the Carathéodory convergence theorem (1912) [**4**, Section 3.1] that the single-slit mappings are dense in S with respect to uniform

convergence on compact subsets of E. A proof of this important property can be found in [**4**, Section 3.2]. Changing it slightly we show that something stronger is true.

Lemma 1. *To each $f \in S$ there corresponds a sequence of single-slit mappings $f_n \in S$, $n = 1, 2, \ldots$, such that $f_n \to f$ uniformly on compact subsets of E as $n \to \infty$, and the boundary of each $f_n(E)$, $n \geq 1$, contains a subray of the negative real axis.*

Proof: Each function $f \in S$ can be approximated uniformly on closed subdisks of E by the functions $r^{-1}f(rz) \in S$, $0 < r < 1$. Thus, it is sufficient to prove our statement for a function $f \in S$ that maps E onto a domain D bounded by an analytic Jordan curve C. In this case there exists a subray L of the negative real axis that belongs to the complement of $\overline{D}$ except for its endpoint $w_L \in C$.

Let J_n be a Jordan arc that runs from infinity along L to the point w_L and then along a portion of C to a point w_n (Figure 1). Let G_n be the complement of J_n and

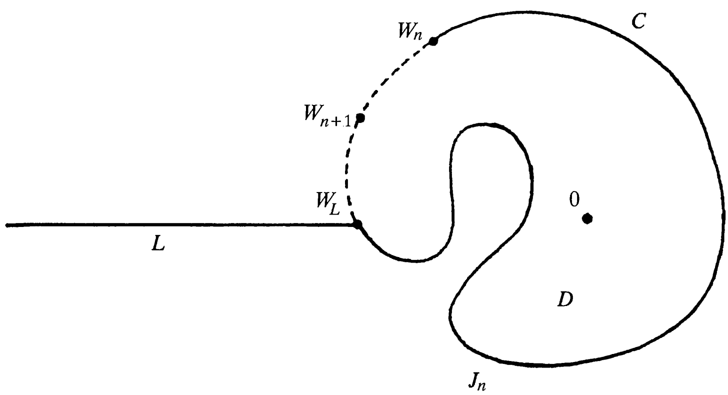


Figure 1

let g_n denote the unique 1–1 analytic map of E onto G_n such that $g_n(0) = 0$ and $g_n'(0) > 0$. Choose the endpoints w_n, $n = 1, 2, \ldots$, so that $J_n \subset J_{n+1}$ and $w_n \to w_L$.

Then D is the component of $\bigcap_{n=1}^{\infty} G_n$ containing the origin. According to the Carathéodory convergence theorem, $g_n \to f$ uniformly on compact subsets of E as $n \to \infty$. Hence $g_n'(0) \to f'(0) = 1$ and we may take $f_n = g_n / g_n'(0)$, $n \geq 1$. ∎

Thus, the single-slit mappings that omit a subray of the negative real axis are dense in S. Similar slit mappings have been considered by G. M. Goluzin, P. P. Kufarev, W. K. Hayman, and other authors. Lemma 1 can be obtained without the Carathéodory theorem, or by just using some of its proof's components. For example, use the Schwarz lemma [**4**, Section 1.1] and the growth theorem (1) to show that the functions g_n, $n \geq 1$, are uniformly bounded on closed subdisks of E. Then it suffices to establish the existence of a subsequence of $\{g_n\}$ that converges uniformly to f on compact sets [**4**; Sections 1.1 and 1.3].

We now state the representation theorem.

Theorem B. (C. Loewner [**8**]). *Let $f \in S$ map E onto the complement of a given Jordan arc $J = \{V(t) : 0 \leq t \leq \infty\}$ (V is 1–1 and continuous) extending from $V(0)$ to infinity. For each $t > 0$ let $f(z, t)$ denote the unique 1–1 analytic map of E onto the*

plane less the portion of J from $V(t)$ to infinity such that $f(0,t)=0$ and $f_z(0,t)>0$, and let $f(z,0)=f(z)$.

The parametrization $V(t)$ can be chosen so that $f_z(0,t)=e^t$, $t>0$. In this case $f(z,t)$ satisfies the partial differential equation

$$f_t(z,t) = zf_z(z,t)\frac{\gamma(t)+z}{\gamma(t)-z}, \qquad z \in E, \quad t \geq 0, \tag{3}$$

where $\gamma(t)$ is a continuous complex-valued function on $[0,\infty)$ with $|\gamma|=1$.

We refer the reader to [**4**; Section 3.3 and Exercise 8, p. 117] for a proof of this classic result. The family $\{f(z,t): t \geq 0\}$ is an example of a so-called *Loewner chain* starting at $f(z)$ and generated by a continuously increasing family of simply connected domains (Figure 2). The point $V(t)$ corresponds to $\gamma(t)$ under the map

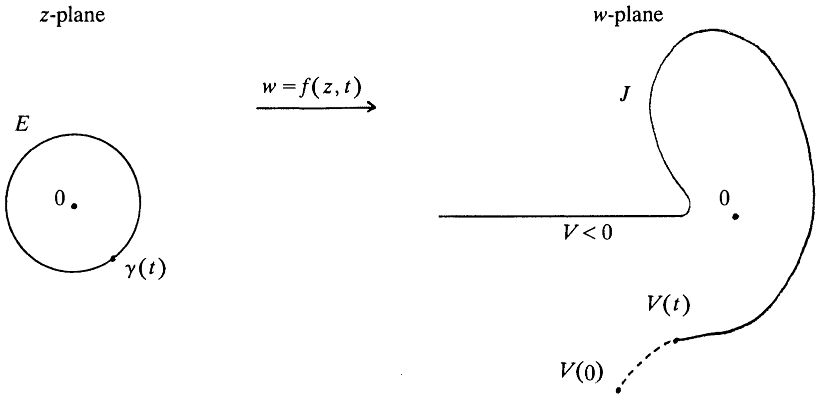


Figure 2

$f(z,t)$. One can use the Schwarz lemma to show that if $V(t)<0$ for $t \geq T \geq 0$ then $f(z,t)=e^tK(z)$ for these values of t. We also note that the Taylor coefficients of $f(z,t)$ are differentiable in the parameter t. This may be shown by using Cauchy's integral formula and differentiation with respect to t under the integral sign.

3. THE MILIN THEOREM AND CONJECTURE: SOMETIMES IT'S BETTER TO ATTACK A HARDER PROBLEM. Since $f(z)/z$ is analytic and zero-free in E for any $f \in S$, we may take $\log[f(z)/z]$ to be the (analytic) branch that vanishes at $z=0$. With this understanding, the logarithmic coefficients of any $f \in S$ are $\{\log[f(z)/z]\}_n$, $n=1,2,\ldots$. For example, $\{\log[K(z)/z]\}_n = 2/n$.

In 1971, I. M. Milin established the following far-reaching connection between the Bieberbach conjecture and the logarithmic coefficients of univalent functions.

Theorem C. (I.M. Milin [**10**, discussion before Theorem 3.2]). *For $f \in S$ and $n \geq 1$, define*

$$I_n(f) \equiv \sum_{m=1}^{n} (n+1-m)\Big(m\left|\{\log[f(z)/z]\}_m\right|^2 - 4/m\Big). \tag{4}$$

If

$$I_n(f) \leq 0 \tag{5}$$

for each $f \in S$ and each $n \geq 1$, then the Bieberbach conjecture is true.

The functionals I_n in (4) are called *Milin's functionals* and (5) is known as *Milin's conjecture*. Since $I_n(K) = 0$ for all $n \geq 1$, Milin's conjecture involves an extremal property of the Koebe function deeper than that of Bieberbach's conjecture. Although the Milin conjecture was verified in certain cases in 1972 [**6**], few seriously believed at the time that one could prove the Bieberbach conjecture through (5). However it was Theorem C that ultimately led researchers out of the dead end they had faced for many years.

Milin's theorem and conjecture go hand in hand with the exponentiation approach developed by Milin in his book [**10**, Chapters 2 and 3]. Properties of formal power series generated by the exponential function play a crucial role in Milin's theory, and the monotonicity lemma [**10**, Lemma 2.2] is the deepest known result of this kind. It has many applications to univalent functions, including a proof of Theorem C. Here we need only a particular case of this lemma, and our proof is similar to the one in [**10**].

Lemma 2. *Let $\{A_m\}_1^\infty$ be an arbitrary sequence of complex numbers, and let the sequence $\{D_m\}_0^\infty$ be defined by the formal expansion*

$$\sum_{m=0}^{\infty} D_m z^m = \exp\left(\sum_{m=1}^{\infty} A_m z^m\right). \tag{6}$$

Let

$$\Theta_n = n^{-1}\sum_{\nu=0}^{n-1} |D_\nu|^2 \exp\left[n^{-1}\sum_{m=1}^{n}(n-m)\left(m^{-1} - m|A_m|^2\right)\right], \quad n \geq 1.$$

Then $\cdots \leq \Theta_n \leq \cdots \leq \Theta_2 \leq \Theta_1 = 1$, and $\Theta_n = 1$ for some $n > 1$ if and only if there is some λ with $|\lambda| = 1$ such that $A_m = \lambda^m/m$, $m = 1, \ldots, n-1$.

Proof: Differentiation of (6) and coefficient comparison yield

$$nD_n = \sum_{m=1}^{n} mA_m D_{n-m}, \quad n \geq 1.$$

By the Cauchy–Schwarz inequality,

$$|D_n|^2 \leq \sigma_n \sum_{m=0}^{n-1} |D_m|^2/n, \quad \text{where} \quad \sigma_n = \sum_{m=1}^{n} |mA_m|^2/n. \tag{7}$$

As

$$\Theta_{n+1}/\Theta_n = \frac{n}{n+1}\left(\sum_{\nu=0}^{n}|D_\nu|^2 \Big/ \sum_{\nu=0}^{n-1}|D_\nu|^2\right)\exp\left[\sum_{m=1}^{n}\left(\frac{1}{n} - \frac{1}{n+1}\right)\left(1 - |mA_m|^2\right)\right],$$

(7) and the inequality $(1-x)e^x \leq 1$ with $x = (1-\sigma_n)/(n+1)$ give

$$\Theta_{n+1}/\Theta_n \leq \frac{n}{n+1}(1 + \sigma_n/n)\exp\left[(1-\sigma_n)/(n+1)\right] \leq 1, \quad n \leq 1.$$

If $\Theta_{n+1} = 1$, then $|A_1|^2 = \sigma_1 = 1$ and there are constants λ_ν ($\nu = 1, \ldots, n$) such that $mA_m = \lambda_\nu \overline{D}_{\nu-m}$, $m = 1, \ldots, \nu$. Since $D_0 = 1$ and $D_1 = A_1$, we get $\nu A_\nu = \lambda_\nu$, $|\lambda_1| = 1$, and $\lambda_\nu = \lambda_{\nu-1}\lambda_1$. It follows that $A_m = \lambda_1^m/m$, $m = 1, \ldots, n$. ■

Proof of Theorem C: Let $f \in S$ and set

$$D_0 = 1, \quad D_m = \left\{[f(z)/z]^{1/2}\right\}_m, \qquad m \geq 1.$$

Then

$$\{f\}_n = \sum_{m=0}^{n-1} D_m D_{n-1-m}, \quad n > 1.$$

The Cauchy-Schwarz inequality gives

$$|\{f\}_n| \leq \sum_{m=0}^{n-1} |D_m|^2. \tag{8}$$

To prove (2) for any given $n > 1$ use (8), the inequality $\Theta_n \leq 1$ from Lemma 2 with $A_m = \{\log[f(z)/z]\}_m/2$ $(m \geq 1)$, and (5) for the preceding value of n.

If $|\{f\}_n| = n$, then Lemma 2 ensures that $\{f\}_2 = \{\log[f(z)/z]\}_1 = 2A_1 = 2\lambda$, where $|\lambda| = 1$. Since the Bieberbach conjecture is already known to be true for $n = 2$ (Theorem A), it follows that $f(z) = \bar{\lambda}K(\lambda z)$. ■

The inequalities $\Theta_n \leq 1$ in Lemma 2 are known as the *Lebedev-Milin exponential inequalities*. In 1967, Milin presented them in a more general form but without proof as a joint work with N. A. Lebedev **[9]**. Milin used his monotonicity lemma to prove them in **[10]**. All published proofs of the Lebedev-Milin inequalities to date are similar to the one in **[10]**.

4. DE BRANGES' DISCOVERY: THE EAGLE HAS LANDED. In 1984, L. de Branges proved a subtle multiparameter inequality for bounded univalent functions that implies the nonpositivity of Milin's functionals (the final version of his proof was published in **[3]**). This result and Milin's Theorem C allowed de Branges to confirm the truth of the Bieberbach conjecture. It came as a

Milin's place, St. Petersburg, Russia (1984). de Branges' proof of Milin's conjecture has been verified. L. de Branges is about to become a mathematical hero. Left to right: seated are Isaak Milin and Louis de Branges; standing are Mrs. Asya Grinshpan, Mrs. Evdokiya Milin, Arcadii Grinshpan, and Evgenii Emelyanov.

sensation to the mathematical world. So far no one has proved the Bieberbach conjecture in any other way, nor has anyone given an essentially different proof of Milin's conjecture.

As a result of discussions that took place during a series of talks given by de Branges to the Goluzin seminar in geometric function theory, de Branges' original proof was verified and reformulated in a simpler form in collaboration with I. M. Milin, E. G. Emelyanov, the author, and others in St. Petersburg, Russia in May, 1984 (see Milin's comments [**11**]). C. FitzGerald and Ch. Pommerenke [**5**], and later L. Weinstein [**12**], further simplified the proof of Milin's conjecture. However their key steps remain the same: apply Loewner's method and use the fact that certain functions introduced by de Branges (de Branges' functions) are nondecreasing in Loewner's parameter t. In general, every de Branges function comes from a Loewner chain starting at some mapping f. In Milin's case it serves as a delicate link between the value of a Milin functional at f and 0 (discussion after Theorem A, Section 2, and (17)).

5. THE AUXILIARY POLYNOMIAL INEQUALITIES: COULD THEY HAVE BEEN PROVED TWO HUNDRED YEARS AGO? The proof that de Branges' functions are nondecreasing is based on the nonnegativity of certain polynomials on the interval $[0,1]$. de Branges recognized this nonnegativity property as a particular case of the Askey-Gasper inequalities for hypergeometric series (1976) [**3**]. Later, Weinstein's simplification [**12**] led to the same property via the addition theorem for Legendre polynomials (1785). See also [**13**], where a mathematician (D. Zeilberger) collaborated with a computer (Shalosh B. Ekhad). More recently, the author and M. E. H. Ismail [**7**] found an elementary and self-contained proof of the polynomial inequalities in question. Here we give a simplified version of this proof.

Lemma 3. *The polynomials $B_{m,n}(x)$ defined by the formal expansion*

$$\left[1-\left(2x^2+(1-x^2)(\zeta+\zeta^{-1})\right)z+z^2\right]^{-1/2} = \sum_{n=0}^{\infty}\left[B_{0,n}(x)+\sum_{m=1}^{n}B_{m,n}(x)(\zeta^m+\zeta^{-m})\right]z^n \qquad (9)$$

can be expressed as

$$B_{m,n}(x) = 4^{-n}\frac{(n-m)!}{(n+m)!}(1-x^2)^m\left[\frac{1}{n!}\left(\frac{d}{dx}\right)^{n+m}(x^2-1)^n\right]^2. \qquad (10)$$

Proof: Taking advantage of an elementary observation used by Th. Clausen (1828), we show that one function is the square of another by means of two linear differential equations.

Given a function $u=u(x)$ and two integers m and n $(0\le m\le n)$, define

$$H[u] \equiv (x^2-1)u''+2(1+m)xu'-(n-m)(n+m+1)u.$$

Coefficient comparison reveals that $u_0(x)=(d/dx)^{n+m}(x^2-1)^n$ satisfies the second-order linear differential equation $H[u]=0$, and hence the equation

$$(x^2-1)u(H[u])'+\left[(2+4m)xu+3(x^2-1)u'\right]H[u]=0. \qquad (11)$$

But we can rewrite (11) as a third-order equation for $h = u^2$ satisfied by $h_0 = u_0^2$:

$$(x^2-1)^2 h''' + 6(1+m)x(x^2-1)h'' + 2\big[(3(2m+1)(m+1) - 2n^2 - 2n)x^2$$
$$-(2m+1)(m+1) + 2n^2 + 2n\big]h' - 4(n-m)(n+m+1)(2m+1)xh = 0. \quad (12)$$

Using Leibniz's rule for the $(n+m)$-th derivative of the product $(x-1)^n(x+1)^n$ we find

$$h_0(1) = u_0^2(1) = 4^{n-m}\left[\frac{n!\,(n+m)!}{m!\,(n-m)!}\right]^2.$$

The next step is to show that each polynomial $B_{m,n}$ differs from a solution of (12) by the m-th power of $(1-x^2)$. Indeed, the left-hand side of (9) can be written as

$$\left[1-(1-x^2)(\zeta^{1/2}-\zeta^{-1/2})^2 z(1-z)^{-2}\right]^{-1/2}\Big/(1-z),$$

so we can use the binomial expansion and the identities

$$(\zeta^{1/2}-\zeta^{-1/2})^{2\nu} = \sum_{m=1}^{\nu}(-1)^{\nu-m}\binom{2\nu}{\nu-m}(\zeta^m+\zeta^{-m}) + (-1)^{\nu}\binom{2\nu}{\nu}, \quad \nu \ge 1,$$

to obtain $B_{m,n}(x) = (1-x^2)^m g(x^2-1)$, where

$$g(x) = \sum_{\nu=0}^{n-m}\frac{x^\nu}{\nu!}\,\frac{(n+m+\nu)!}{(n-m-\nu)!\,(\nu+2m)!}\,\frac{(1/2)_{\nu+m}}{(\nu+m)!}.$$

The symbol $(1/2)_k$ denotes the *shifted factorial* of $1/2$:

$$(1/2)_0 = 1 \quad\text{and}\quad (1/2)_k = (1/2)(1/2+1)\cdots(1/2+k-1), \quad k = 1,2,\ldots.$$

To see that $G(x) = g(x^2-1)$ satisfies (12), substitute $G(x)$ for $h(x)$ in the left-hand side of (12), divide through by x, and consider the result as a polynomial in (x^2-1). It is identically zero if and only if all of its coefficients are zero, i.e., if and only if

$$2\nu(\nu-1)(\nu-2) + 3\nu(\nu-1) + 6(1+m)\nu(\nu-1) + 3(1+m)\nu$$
$$+\big[3(2m+1)(m+1) - 2n^2 - 2n\big]\nu - (2m+1)(n-m)(n+m+1)$$
$$+\big[\nu(\nu-1) + 3(1+m)\nu + (2m+1)(m+1)\big](n-m-\nu) \quad (13)$$
$$\cdot(2\nu+2m+1)(n+\nu+m+1)/\big[(\nu+2m+1)(\nu+m+1)\big] = 0$$

for fixed n and m and corresponding values of the index ν. Since

$$\nu(\nu-1) + 3(1+m)\nu + (2m+1)(m+1) = (\nu+2m+1)(\nu+m+1),$$

(13) is easy to verify.

Finally, we observe that a polynomial solution of (12) is uniquely determined by its value at $x = 1$. Indeed, we can write this solution as a polynomial in $(x-1)$. Then (12) allows us to find all of its coefficients step-by-step. Because G and h_0 are polynomials and

$$G(1) = \frac{(n+m)!}{(n-m)!}\,\frac{(1/2)_m}{(2m)!\,m!} = \frac{4^{-n}}{(n!)^2}\,\frac{(n-m)!}{(n+m)!}h_0(1),$$

(10) follows. ∎

Corollary 1. *The polynomials* $P_{m,n}(x)$ *defined by the formal expansion*

$$\left[1-\left(2(1-x)+x(\zeta+\zeta^{-1})\right)z+z^2\right]^{-1}$$
$$=\sum_{n=0}^{\infty}\left[\sum_{m=0}^{n}P_{m,n}(x)(\zeta^m+\zeta^{-m})\right]z^n \tag{14}$$

are nonnegative for $x\in[0,1]$.

6. DE BRANGES' THEOREM (THE PROOF OF MILIN'S CONJECTURE): A FUNCTION WITH NONNEGATIVE DERIVATIVE IS NONDECREASING.

In this section we combine the main idea of de Branges with a coefficient representation of de Branges' functions to present a short proof of Milin's conjecture. We take into account some observations from **[5]** and **[12]** on treating Loewner's equation and the auxiliary polynomials.

Theorem D. (L. de Branges **[3]**). *For each* $f\in S$ *and each* $n\geq 1$, *Milin's functionals* (4) *satisfy*

$$I_n(f)\leq 0. \tag{15}$$

To prove Theorem D we need one more lemma.

Lemma 4. *Let* a_m, $m=1,2,\dots,$ *be given and define* $b_m=2(1+\Sigma_{\nu=1}^{m}a_\nu)-a_m$, $m=1,2,\dots$. *Then*

$$4\mathrm{Re}\left(1+\sum_{\nu=1}^{m}\bar{a}_\nu b_\nu\right)=|a_m+b_m|^2,\quad m\geq 1. \tag{16}$$

Proof: We use induction on m. Since $a_k+b_k=b_{k+1}-a_{k+1}$, (16) holds for $m=(k+1)$ if it is valid for $m=k$, and (16) holds for $m=1$ because $b_1-a_1=2$. ∎

Proof of Theorem D: Fix a natural number n.

Since I_n is a continuous functional on S [**4**, Section 1.4], it is sufficient to prove (15) for the dense subclass of S consisting of all single-slit mappings that omit a subray of the negative real axis (Lemma 1). Fix $f(z)$ in this subclass and construct the Loewner chain $\{f(z,t): t\geq 0\}$ as in Theorem B. Then there exists some $T\geq 0$ such that $f(z,t)=e^tK(z)$ for $t\geq T$ (discussion after Theorem B). Define the differentiable function

$$\varphi_n(t)=\left\{K(z)\sum_{m=1}^{n}\left(m|c_m(t)|^2-4/m\right)w^m(z,t)\right\}_{n+1},\quad t\in[0,T], \tag{17}$$

where $c_m(t)=\{\log[f(z,t)/z]\}_m$ and $w(z,t)$ is the Pick function defined implicitly by the equation

$$e^tK(w(z,t))=K(z),\qquad z\in E,\quad t\geq 0.$$

Clearly, $w(z,0)=z$ and hence $\varphi_n(0)=I_n(f)$. Also $\varphi_n(T)=0$ since $c_m(T)=\{\log[e^TK(z)/z]\}_m=2/m$, $m\geq 1$. Thus it is enough to show that $\varphi_n'\geq 0$.

First we use the definitions of $w = w(z,t)$ and $K(z)$ to get

$$K(z)\frac{1-w}{1+w}\left[1+\sum_{m=1}^{\infty} w^m(\zeta^m+\zeta^{-m})\right]$$

$$= z\left[1-\left(2(1-e^{-t})+e^{-t}(\zeta+\zeta^{-1})\right)z+z^2\right]^{-1},$$

provided $\zeta \neq 0$ and $|w| < |\zeta|, |\zeta|^{-1}$. This equation and (14) give

$$P_{m,n}(e^{-t}) = \left\{K(z)\frac{1-w(z,t)}{1+w(z,t)}w^m(z,t)\right\}_{n+1}, \qquad 1 \leq m \leq n. \tag{18}$$

Now we observe that $w(z,t)$ satisfies the equation $w_t/w = (w-1)/(w+1)$, and use the expansion $(1+w)/(1-w) = 1 + 2w + 2w^2 + \cdots$ and (18) to compute φ_n':

$$\begin{aligned}
\varphi_n'(t) &= \frac{d}{dt}\left\{K(z)\left[\sum_{m=1}^{n}\left(m|c_m(t)|^2-4/m\right)w^m(z,t)\right]\right\}_{n+1}\\
&= \left\{K(z)\sum_{m=1}^{n}\left[2\operatorname{Re}\left(mc_m'(t)\overline{c_m(t)}\right)\right.\right.\\
&\qquad\left.\left.+\left(|mc_m(t)|^2-4\right)\frac{w(z,t)-1}{w(z,t)+1}\right]w^m(z,t)\right\}_{n+1}\\
&= \sum_{m=1}^{n} P_{m,n}(e^{-t})\operatorname{Re}\left[4\left(1+\sum_{\nu=1}^{m}\nu c_\nu'(t)\overline{c_\nu(t)}\right)-2mc_m'(t)\overline{c_m(t)}-|mc_m(t)|^2\right].
\end{aligned} \tag{19}$$

Dividing (3) by $f(z,t)$ and comparing the coefficients of z^m on both sides, we find

$$c_m'\gamma^m = 2\left(1+\sum_{\nu=1}^{m}\nu c_\nu\gamma^\nu\right) - mc_m\gamma^m, \quad m \geq 1.$$

We are now in a position to apply Lemma 4 with $a_m = mc_m\gamma^m$ and $b_m = c_m'\gamma^m$, $m = 1,\ldots,n$, to the last line in (19). We have

$$\varphi_n'(t) = \sum_{m=1}^{n} P_{m,n}(e^{-t})|c_m'(t)|^2. \tag{20}$$

Corollary 1 and (20) show that $\varphi_n'(t) \geq 0$, $t \in [0,T]$. It follows that $I_n(f) \leq 0$. ∎

Although our proof did not require a geometric description of the Pick function, we mention that for each $t > 0$, $w(z,t) = e^{-t}z + \cdots$ maps E onto E cut along the negative real axis from -1 to $1 + 2[(e^{2t} - e^t)^{1/2} - e^t]$.

7. PROOF OF THE BIEBERBACH CONJECTURE. Theorems C and D imply that the Bieberbach conjecture is true:

Main Theorem. (L. de Branges [3]). *Let $f \in S$. Then*

$$|\{f\}_n| \leq n, \quad n = 2,3,\ldots.$$

Equality holds for any given n only for the Koebe function $K(z) = z/(1-z)^2$ and its rotations $\bar{\lambda}K(\lambda z)$, $|\lambda| = 1$.

REFERENCES

1. *The Bieberbach Conjecture, Proceedings of the Symposium on the Occasion of the Proof*, Editors: A. Baernstein II, D. Drasin, P. Duren, and A. Marden, Amer. Math. Soc., Providence, RI, 1986.
2. L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, *Sitzungsber. Preuss. Akad. Wiss.* **38** (1916) 940–955.
3. L. de Branges, A proof of the Bieberbach conjecture, *Acta Math.* **154** (1985) 137–152.
4. P. L. Duren, *Univalent Functions*, Springer-Verlag, Heidelberg and New York, 1983.
5. C. FitzGerald and Ch. Pommerenke, The de Branges theorem on univalent functions, *Trans. Amer. Math. Soc.* **290** (1985) 683–690.
6. A. Z. Grinshpan, Logarithmic coefficients of functions in the class *S*, *Sibirsk. Mat. Z.* **13** (1972) 1145–1157 (in Russian); *Siberian Math. J.* **13** (1972) 793–801.
7. A. Z. Grinshpan and M. E. H. Ismail, On a sequence of polynomials related to the Bieberbach – Milin problem, *Abstracts Amer. Math. Soc.* **17** (1996) 61.
8. K. Löwner (C. Loewner), Untersuchungen über schlichte konforme Abbildungen des Einheitskreises. I, *Math. Ann.* **89** (1923) 103–121.
9. I. M. Milin, On the coefficients of univalent functions, *Dokl. Akad. Nauk SSSR* **176** (1967) 1015–1018 (in Russian); *Soviet Math. Dokl.* **8** (1967) 1255–1258.
10. I. M. Milin, *Univalent Functions and Orthonormal Systems*, Nauka, Moscow, 1971 (in Russian); published in English by Amer. Math. Soc., Providence, RI, 1977.
11. I. M. Milin, Comments on the proof of the conjecture on logarithmic coefficients, in [**1**], pp. 109–112.
12. L. Weinstein, The Bieberbach conjecture, *Internat. Math. Res. Notices* **5** (1991) 61–64.
13. S. B. Ekhad and D. Zeilberger, A high-school algebra, "formal calculus," proof of the Bieberbach conjecture [after L. Weinstein], *Contemp. Math.* **178** (1994) 113–115.

ARCADII Z. GRINSHPAN was born (1945) and educated (Ph.D., 1973) in St. Petersburg, Russia. Since his student years at St. Petersburg University, he has been engaged in both pure and applied mathematical research: geometric function theory, inequalities, mathematical modeling in diesel engineering, and numerical analysis. Now his experience is used to teach future mathematicians and engineers at the University of South Florida.
University of South Florida, Tampa, FL 33620
azg@math.usf.edu

From the MONTHLY 100 years ago...

Harvard University has just issued its Course of Study in Mathematics for the year 1899–1900. Among the courses offered in advanced Mathematics are the following: General Theory of Surface, Prof. J. M. Peirce; Dynamics of a Rigid Body, Professor Byerly; Quaternions with Applications to Geometry and Mechanics, Prof. J. M. Peirce; Trigonometric series, Introduction to Spherical Harmonics, Potential Functions, Professors Byerly and B. O. Peirce; Theory of Functions (Second Course), Riemann's Theory of Functions, Professor Osgood; Algebra—Galois's Theory of Equations, Professor Osgood; Lie's Theories as Applied to Differential Equations, Dr. Bouton; etc., etc. The courses offered at Harvard are sufficiently varied and extensive to meet the wants of any student of mathematics.

...Vol. 6 (1899) 159

Decision Making: A Golden Rule

Dimitris A. Sardelis and Theodoros M. Valahas

1. INTRODUCTION. Suppose someone suggests to you the following game: You are able to take as many slips of paper as you please and on each slip write a different number without restrictions. Then, you turn the slips face down, shuffle them, and start turning them face up, one at a time. As the numbers present themselves one after the other, the proponent of the game is to interrupt their procession by speculating that some number just passed is the largest of the sequence. He is to make a single guess about a number being the largest right at the moment that it shows up. If all slips have been turned over and he has not yet pronounced a preference, he must "choose" the last number. The proponent of the game is courteous enough to play the game with any odds you consider to be fair. What would you suggest?

This problem first appeared in the February 1960 issue of *Scientific American* [**7**]. Since then, it has been extended [**1**–**2**] and generalized [**3**–**9**] in many directions by eminent probabilists and statisticians so one may justly claim that it now constitutes a distinct field of study within probability-optimization theories. It has come to be known as the beauty contest problem, the secretary problem, the marriage problem or fiancé problem, and the dowry problem.

The importance of the game is that it provides an artificial-idealized simulation of sequential decision processes. Indeed, everyday life reveals that almost all successful decisions are preceded by a learning period during which one observes, classifies, and ranks experiences. Given a finite life-span for some decision making, many alternate strategies can be pursued, every one of which is specified by the ratio between learning and acting-decision periods one agrees to employ. The solution offered by the idealized mathematical version of the problem is remarkable: The optimum strategy is attained when about e^{-1} of the available decision time is devoted to learning. The probability of success for this optimum strategy is also about e^{-1}, which is approximately 37%. This simple and elegant rule rightly deserves to be called *a golden rule for decision making.*

We explore the problem and construct its solution through an ongoing, developing learning process. At first, we elaborate on the alternative possible strategies by listing and enumerating the cases where particular strategies win. Thus, we form a preliminary conception of the optimization nature of the problem and we derive a general expression for a strategy's probability of success. This expression is then used to determine a probability spectrum for the winning strategies in cases where direct listing and counting are not possible. A pattern of optimal strategies emerges, ultimately expressed by general conditions. These in turn yield a further expansion of the horizons of the problem that culminates in the devising of (a) a practical guide for making optimal decisions and (b) a very efficient rule for estimating the decision span for which any particular strategy becomes optimal. Finally, exploration of the optimization conditions for the winning strategies leads to the golden rule for decisions.

2. PROBLEM DESCRIPTION. The original problem with the slips of paper may be restated more formally as follows: A known number N of items is to be

presented to an observer one by one in random order, all possible orderings being equally likely. The observer is able at any time to rank without ties the items that have so far been presented in order of desirability. As each item is presented he must either accept it, in which case the process stops, or reject it, in which case the next item in the sequence is presented and the observer faces the same dilemma as before. If the last item is ever reached it must be accepted. The observer's aim is to find the best of the N items available by employing a strategy with as high a probability of success as possible.

3. ON THE POSSIBLE STRATEGIES. The observer must either accept or reject an item right at the moment that it is presented, i.e., he cannot go back and choose an already-presented item that, in retrospect, turns out to be best. He has to balance the risk of stopping too soon and accepting an apparently desirable item when an even better one might be still to come, against that of going on too long and discovering that the best item was rejected earlier.

All possible strategies range between two equally likely extremes that constitute the worst choices an observer can make. On the one hand, an observer might pick the first item. On the other hand, he might wait for the last item. The probability of success for both these trivial strategies is the same and equals $1/N$.

Consequently, getting some experience from the contest process—by observing, comparing, and ranking items—before reaching a decision, cannot make things worse. On the contrary, there is hope for improving one's chances. Let us define the following $N-2$ non trivial S_n strategies:

> The observer lets n items pass, $1 \le n \le N-2$, ranks them in order of desirability, and then among the next items selects the first one found with a higher rank.

Among all possible strategies S_n, the observer wants to select and employ the one with the maximum probability of success.

4. EXPLICIT SOLUTIONS. To gain some insight into the problem context, it is instructive to explore the simplest cases first, i.e., when the total number N of items is small.

Let us evaluate the probability of success for each S_n strategy explicitly when $N = 3, 4, 5$. Our goal will be achieved by brute force, i.e., by listing in every such case all possible orderings and their associated winning strategies (the items are represented by their ranks). The optimal strategy in all cases will be the one that wins most often.

TABLE 1. Orderings and Winning Strategies ($N = 3$)

1	1	2	3	
2	1	3	2	S_1
3	2	1	3	S_1
4	2	3	1	S_1
5	3	1	2	
6	3	2	1	

Case $N = 3$ There are 6 possible orderings $(\mathbf{1}, \mathbf{2}, \ldots, \mathbf{6})$ and one non-trivial strategy, S_1 (see Table 1). Thus S_1 wins in the orderings **2**, **3**, **4** and loses in all others. Therefore, the probability of success for S_1 is $3/6 = 50\%$.

Case $N = 4$ There are 24 orderings in this case and there are two non-trivial strategies: S_1, S_2 (see Table 2). Strategies S_1 and S_2, are successful with probabilities $P(S_1) = 11/24$ and $P(S_2) = 10/24$. Therefore, S_1 is the optimal strategy.

TABLE 2. Orderings and Winning Strategies ($N = 4$)

1	1	2	3	4		**13**	3	1	2	4	S_1 S_2
2	1	2	4	3	S_2	**14**	3	1	4	2	S_1 S_2
3	1	3	2	4	S_2	**15**	3	2	1	4	S_1 S_2
4	1	3	4	2	S_2	**16**	3	2	4	1	S_1 S_2
5	1	4	2	3	S_1	**17**	3	4	1	2	S_1
6	1	4	3	2	S_1	**18**	3	4	2	1	S_1
7	2	1	3	4		**19**	4	1	2	3	
8	2	1	4	3	S_1 S_2	**20**	4	1	3	2	
9	2	3	1	4	S_2	**21**	4	2	1	3	
10	2	3	4	1	S_2	**22**	4	2	3	1	
11	2	4	1	3	S_1	**23**	4	3	1	2	
12	2	4	3	1	S_1	**24**	4	3	2	1	

Case $N = 5$ There are 120 orderings here and there are three non-trivial strategies: S_1, S_2, and S_3 with probabilities of success $P(S_1) = 50/120$, $P(S_2) = 52/120$, and $P(S_3) = 42/120$ (see Table 3). Therefore, the optimal strategy is S_2.

After this direct listing and enumeration of cases, we see that

(i) the non-trivial strategies do indeed improve odds compared to a chance selection, and
(ii) among these strategies, some are better than others.

5. A STRATEGY'S PROBABILITY OF SUCCESS. The search for good strategies by listing of all orderings becomes a forbidding task for larger N. Even for $N = 10$ there are about 3.5 million, while for $N = 15$ there are one trillion! Consequently, we must figure out some abstract way to estimate the probability of success for strategies.

Let us start with some general observations. A strategy S_n may fail in two ways:

- The best item may be included in the n items defining S_n. As an example, for $N = 5$ we have three strategies: S_1, S_2, and S_3. Strategy S_3 loses whenever number 5 appears first or second or third, S_2 loses whenever number 5 appears first or second and, finally, S_1 loses whenever number 5 appears first.
- The best item may be preceded by at least one item whose rank exceeds those of the first n items. For example, when $N = 4$ strategy S_2 loses in orderings **1** and **7** where one chooses as best number 3 instead of 4. Similarly, for $N = 5$ strategy S_2 loses in 68 orderings, (e.g., **26** and **52**), while S_3 loses in 78 orderings (e.g., **31** and **55**).

TABLE 3. Orderings and Winning Strategies ($N = 5$)

1	1	2	3	4	5			
2	1	2	3	5	4			S_3
3	1	2	4	3	5			S_3
4	1	2	4	5	3			S_3
5	1	2	5	3	4		S_2	
6	1	2	5	4	3		S_2	
7	1	3	2	4	5			
8	1	3	2	5	4		S_2	S_3
9	1	3	4	2	5			S_3
10	1	3	4	5	2			S_3
11	1	3	5	2	4		S_2	
12	1	3	5	4	2		S_2	
13	1	4	2	3	5		S_2	S_3
14	1	4	2	5	3		S_2	S_3
15	1	4	3	2	5		S_2	S_3
16	1	4	3	5	2		S_2	S_3
17	1	4	5	2	3		S_2	
18	1	4	5	3	2		S_2	
19	1	5	2	3	4	S_1		
20	1	5	2	4	3	S_1		
21	1	5	3	2	4	S_1		
22	1	5	3	4	2	S_1		
23	1	5	4	2	3	S_1		
24	1	5	4	3	2	S_1		
25	2	1	3	4	5			
26	2	1	3	5	4			S_3
27	2	1	4	3	5			S_3
28	2	1	4	5	3			S_3
29	2	1	5	3	4	S_1	S_2	
30	2	1	5	4	3	S_1	S_2	
31	2	3	1	4	5			
32	2	3	1	5	4		S_2	S_3
33	2	3	4	1	5			S_3
34	2	3	4	5	1			S_3
35	2	3	5	1	4		S_2	
36	2	3	5	4	1		S_2	
37	2	4	1	3	5		S_2	S_3
38	2	4	1	5	3		S_2	S_3
39	2	4	3	1	5		S_2	S_3
40	2	4	3	5	1		S_2	S_3
41	2	4	5	1	3		S_2	
42	2	4	5	3	1		S_2	
43	2	5	1	3	4	S_1		
44	2	5	1	4	3	S_1		
45	2	5	3	1	4	S_1		
46	2	5	3	4	1	S_1		
47	2	5	4	1	3	S_1		
48	2	5	4	3	1	S_1		
49	3	1	2	4	5			
50	3	1	2	5	4	S_1	S_2	S_3
51	3	1	4	2	5			S_3
52	3	1	4	5	2			S_3
53	3	1	5	2	4	S_1	S_2	
54	3	1	5	4	2	S_1	S_2	
55	3	2	1	4	5			
56	3	2	1	5	4	S_1	S_2	S_3
57	3	2	4	1	5			S_3
58	3	2	4	5	1			S_3
59	3	2	5	1	4	S_1	S_2	
60	3	2	5	4	1	S_1	S_2	
61	3	4	1	2	5		S_2	S_3
62	3	4	1	5	2		S_2	S_3
63	3	4	2	1	5		S_2	S_3
64	3	4	2	5	1		S_2	S_3
65	3	4	5	1	2		S_2	
66	3	4	5	2	1		S_2	
67	3	5	1	2	4	S_1		
68	3	5	1	4	2	S_1		
69	3	5	2	1	4	S_1		
70	3	5	2	4	1	S_1		
71	3	5	4	1	2	S_1		
72	3	5	4	2	1	S_1		
73	4	1	2	3	5	S_1	S_2	S_3
74	4	1	2	5	3	S_1	S_2	S_3
75	4	1	3	2	5	S_1	S_2	S_3
76	4	1	3	5	2	S_1	S_2	S_3
77	4	1	5	2	3	S_1	S_2	
78	4	1	5	3	2	S_1	S_2	
79	4	2	1	3	5	S_1	S_2	S_3
80	4	2	1	5	3	S_1	S_2	S_3
81	4	2	3	1	5	S_1	S_2	S_3
82	4	2	3	5	1	S_1	S_2	S_3
83	4	2	5	1	3	S_1	S_2	
84	4	2	5	3	1	S_1	S_2	
85	4	3	1	2	5	S_1	S_2	S_3
86	4	3	1	5	2	S_1	S_2	S_3
87	4	3	2	1	5	S_1	S_2	S_3
88	4	3	2	5	1	S_1	S_2	S_3
89	4	3	5	1	2	S_1	S_2	
90	4	3	5	2	1	S_1	S_2	
91	4	5	1	2	3	S_1		
92	4	5	1	3	2	S_1		
93	4	5	2	1	3	S_1		
94	4	5	2	3	1	S_1		
95	4	5	3	1	2	S_1		
96	4	5	3	2	1	S_1		
97	5	1	2	3	4			
98	5	1	2	4	3			
99	5	1	3	2	4			
100	5	1	3	4	2			
101	5	1	4	2	3			
102	5	1	4	3	2			
103	5	2	1	3	4			
104	5	2	1	4	3			
105	5	2	3	1	4			
106	5	2	3	4	1			
107	5	2	4	1	3			
108	5	2	4	3	1			
109	5	3	1	2	4			
110	5	3	1	4	2			
111	5	3	2	1	4			
112	5	3	2	4	1			
113	5	3	4	1	2			
114	5	3	4	2	1			
115	5	4	1	2	3			
116	5	4	1	3	2			
117	5	4	2	1	3			
118	5	4	2	3	1			
119	5	4	3	1	2			
120	5	4	3	2	1			

Consequently, S_n is a winning strategy if

(a) *the best item is a candidate for selection, and*
(b) *the ranks of items, if any, preceding the best, do not exceed those of the first n items.*

These two conditions lead to a general expression for the probability of success of strategy S_n, henceforth denoted as $P_N(S_n)$.

Let E_k denote the event that the best item is at some position k, i.e., it is the kth term in the N-item sequence. Since all orderings are assumed equally likely, then by condition (a), the respective probability is $P(E_k) = 1/N$ with $k > n$.

Let F_k denote the event described by condition (b). Then the probability that F_k occurs, i.e., that the highest rank of the first $k-1$ terms appears in the first n terms, is $P(F_k) = n/(k-1)$.

S_n is a winning strategy if both conditions (a) and (b) are satisfied. Since events E_k and F_k are independent, and all E_k events are exclusive and exhaustive alternatives, we have

$$\begin{aligned} P_N(S_n) &= \sum_{k=n+1}^{N} P(E_k \cap F_k) \\ &= \sum_{k=n+1}^{N} P(E_k) \cdot P(F_k) = \frac{n}{N} \cdot \sum_{k=n+1}^{N} \left(\frac{1}{k-1} \right) \end{aligned} \tag{1}$$

This is the desired expression for the probability of success for any strategy S_n and for any number N of items one cares to choose from. The $P_N(S_n)$ expression is also valid for $n = N - 1$, in which case it equals

$$P_N(S_{N-1}) = [(N-1)/N] \cdot [1/(N-1)] = 1/N,$$

as it should be.

Since we have made no assumptions about the distribution of items, we safely say that the S_n decision rule is general.

6. THE PROBABILITY SPECTRUM FOR THE WINNING STRATEGIES. The $P_N(S_n)$ general expression in (1) provides a powerful tool for evaluating the probabilities of winning for all S_n strategies without actually having to list the possible orderings and count the corresponding winning frequencies. Table 4 exemplifies this point for $N = 1, 2, \ldots, 20$ and all possible strategies S_n, $0 \leq n \leq N - 1$. Every entry is the probability of success of a strategy S_n when the number of items is N, expresed with four significant digits.

7. OPTIMAL STRATEGIES. Each column in Table 4 possesses a maximum probability for some value of N. For example, in columns S_4 and S_7, the probabilities of success are 0.3984 for $N = 11$ and 0.3850 for $N = 19$. Furthermore, we see that every row corresponding to a particular N-value possesses a maximum probability of success for some strategy. For example, the maximum probabilities of success for $N = 10$ and $N = 16$ are 0.3987 and 0.3881, respectively, and they correspond to strategies S_3 and S_6, respectively.

The emerging pattern of optimal strategies may be expressed by the following statements:

(a) *To every number N of items there corresponds one strategy S_n with a maximum probability of success, and conversely,*
(b) *every possible strategy S_n is optimal for a particular number N of items.*

TABLE 4. Probabilities of Success for Strategies (N = 1 to 20)

N	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}
1	1.0000																			
2	0.5000	0.5000																		
3	0.3333	**0.5000**	0.3333																	
4	0.2500	0.4583	0.4167	0.2500																
5	0.2000	0.4167	**0.4333**	0.3500	0.2000															
6	0.1667	0.3806	0.4278	0.3917	0.3000	0.1667														
7	0.1429	0.3500	0.4143	0.4071	0.3524	0.2619	0.1429													
8	0.1250	0.3241	0.3982	**0.4098**	0.3798	0.3185	0.2321	0.1250												
9	0.1111	0.3020	0.3817	0.4060	0.3931	0.3525	0.2897	0.2083	0.1111											
10	0.1000	0.2829	0.3658	0.3987	0.3983	0.3728	0.3274	0.2653	0.1889	0.1000										
11	0.0909	0.2663	0.3507	0.3897	**0.3984**	0.3844	0.3522	0.3048	0.2444	0.1727	0.0909									
12	0.0833	0.2517	0.3366	0.3800	0.3955	0.3902	0.3683	0.3324	0.2847	0.2265	0.1591	0.0833								
13	0.0769	0.2387	0.3236	0.3700	0.3907	**0.3923**	0.3784	0.3517	0.3141	0.2668	0.2110	0.1474	0.0769							
14	0.0714	0.2272	0.3114	0.3600	0.3848	0.3917	0.3843	0.3651	0.3356	0.2972	0.2508	0.1973	0.1374	0.0714						
15	0.0667	0.2168	0.3002	0.3503	0.3782	0.3894	0.3873	0.3741	0.3513	0.3202	0.2817	0.2366	0.1853	0.1286	0.0667					
16	0.0625	0.2074	0.2898	0.3409	0.3712	0.3859	**0.3881**	0.3799	0.3627	0.3377	0.3058	0.2676	0.2238	0.1747	0.1208	0.0625				
17	0.0588	0.1989	0.2801	0.3319	0.3641	0.3816	0.3873	0.3832	0.3708	0.3509	0.3246	0.2923	0.2547	0.2122	0.1652	0.1140	0.0588			
18	0.0556	0.1911	0.2711	0.3233	0.3569	0.3767	0.3854	0.3848	0.3763	0.3608	0.3392	0.3120	0.2798	0.2429	0.2018	0.1567	0.1078	0.0556		
19	0.0526	0.1840	0.2626	0.3150	0.3498	0.3715	0.3827	**0.3850**	0.3799	0.3682	0.3506	0.3278	0.3001	0.2681	0.2321	0.1923	0.1490	0.1023	0.0526	
20	0.0500	0.1774	0.2548	0.3072	0.3429	0.3661	0.3793	0.3842	0.3820	0.3734	0.3594	0.3403	0.3167	0.2889	0.2573	0.2221	0.1836	0.1420	0.0974	0.0500

For fixed N, the winning probabilities for any two successive strategies, S_n and S_{n+1}, differ by

$$\Delta P_{(n)} = P_N(S_{n+1}) - P_N(S_n) = \frac{1}{N}\left[\sum_{k=n+2}^{N}\left(\frac{1}{k-1}\right) - 1\right]. \tag{2}$$

The optimal strategy corresponds to the smallest n that makes $\Delta P_{(n)}$ negative. Therefore, the best n for a given fixed N is the least n such that

$$\sum_{k=n+2}^{N}\left(\frac{1}{k-1}\right) = \frac{1}{n+1} + \frac{1}{n+2} + \frac{1}{n+3} + \cdots + \frac{1}{N-1} < 1. \tag{3}$$

For fixed n, the winning probabilities of S_n for any two consecutive N-values, N and $N+1$, differ by

$$\Delta P_{(N)} = P_{N+1}(S_n) - P_N(S_n) = \frac{n}{N(N+1)}\left[1 - \sum_{k=n+1}^{N}\left(\frac{1}{k-1}\right)\right]. \tag{4}$$

It follows that S_n is best for the smallest N-value that makes $\Delta P_{(N)}$ negative. Therefore, the best N for a given fixed strategy S_n is the least N such that

$$\sum_{k=n+1}^{N}\left(\frac{1}{k-1}\right) = \frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{N-1} > 1. \tag{5}$$

The latter condition expands our computational and conceptual horizons of the problem considerably. Table 5 presents the N-values where specific S_n strategies (fixed n) become most appropriate, i.e., they attain the maximum possible probability of success. This probability is also maximum when all possible S_n strategies are compared for the same N.

8. GEOMETRIC EVALUATION OF PROBABILITY FOR OPTIMAL STRATEGIES. The treatment of the problem may be extended still more with quite interesting results. In what follows we shall consider condition (5) for optimal strategies within the realm of the real number continuum. Consequently, we shall deduce an efficient rule for identifying the best N-value for any S_n strategy.

Every term of the sum in (5), say the term $1/m$, may be represented as the area of the rectangle $m_-M_-M_+m_+$ (see Figure 1) with base $(m_-m_+) = 1$ and height $(Mm) = 1/m$. Evidently, M is a point on the hyperbola $y = 1/x$. The area (LM_-M) below the hyperbola is

$$E \equiv (LM_-M) = \int_{m-1/2}^{m}\left(\frac{1}{x} - \frac{1}{m}\right)dx = \ln\left(\frac{2m}{2m-1}\right) - \frac{1}{2m}. \tag{6}$$

Similarly, the area (MM_+N) above the hyperbola is

$$\varepsilon \equiv (MM_+N) = \int_{m}^{m+1/2}\left(\frac{1}{m} - \frac{1}{x}\right)dx = \frac{1}{2m} - \ln\left(\frac{2m+1}{2m}\right). \tag{7}$$

TABLE 5. A Practical Guide on Making the Optimal Decision when Choosing the Best out of a Number N of Contestants Presenting Themselves in Sequence

n	N	Pr (%)	n	N	Pr (%)	n	N	Pr (%)	n	N	Pr (%)
1	**3**	50.00	26	**70**	37.24	51	**138**	37.02	76	**206**	36.94
2	**5**	43.33	27	**73**	37.22	52	**141**	37.01	77	**209**	36.94
3	**8**	40.98	28	**76**	37.21	53	**144**	37.01	78	**212**	36.94
4	**11**	39.84	29	**78**	37.19	54	**146**	37.00	79	**214**	36.94
5	**13**	39.23	30	**81**	37.18	55	**149**	37.00	80	**217**	36.93
6	**16**	38.81	31	**84**	37.17	56	**152**	37.00	81	**220**	36.93
7	**19**	38.50	32	**87**	37.15	57	**155**	36.99	82	**223**	36.93
8	**21**	38.28	33	**89**	37.14	58	**157**	36.99	83	**225**	36.93
9	**24**	38.12	34	**92**	37.13	59	**160**	36.99	84	**228**	36.93
10	**27**	37.98	35	**95**	37.12	60	**163**	36.98	85	**231**	36.93
11	**30**	37.87	36	**98**	37.11	61	**165**	36.98	86	**233**	36.92
12	**32**	37.78	37	**100**	37.10	62	**168**	36.98	87	**236**	36.92
13	**35**	37.70	38	**103**	37.10	63	**171**	36.97	88	**239**	36.92
14	**38**	37.63	39	**106**	37.09	64	**174**	36.97	89	**242**	36.92
15	**40**	37.57	40	**108**	37.08	65	**176**	36.97	90	**244**	36.92
16	**43**	37.53	41	**111**	37.07	66	**179**	36.96	91	**247**	36.92
17	**46**	37.48	42	**114**	37.07	67	**182**	36.96	92	**250**	36.91
18	**49**	37.44	43	**117**	37.06	68	**184**	36.96	93	**252**	36.91
19	**51**	37.41	44	**119**	37.05	69	**187**	36.96	94	**255**	36.91
20	**54**	37.38	45	**122**	37.05	70	**190**	36.95	95	**258**	36.91
21	**57**	37.35	46	**125**	37.04	71	**193**	36.95	96	**261**	36.91
22	**59**	37.32	47	**127**	37.04	72	**195**	36.95	97	**263**	36.91
23	**62**	37.30	48	**130**	37.03	73	**198**	36.95	98	**266**	36.91
24	**65**	37.28	49	**133**	37.03	74	**201**	36.95	99	**269**	36.91
25	**68**	37.26	50	**136**	37.02	75	**204**	36.94	100	**271**	36.90

Decision Rule: *Let n of out N contestants pass, rank them in order of desirability, and then choose the first contestants who rank higher.*

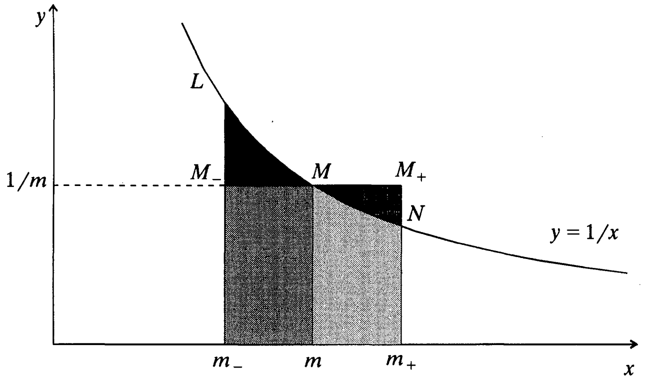


Figure 1

Consequently, the area of the region $LM_{-}M$ is larger than the area of the region $MM_{+}N$, i.e., $E > \varepsilon$, since

$$E - \varepsilon = \ln\left(\frac{2m+1}{2m-1}\right) - \frac{1}{m} = 2\left[\frac{1}{3(2m)^3} + \frac{1}{5(2m)^5} + \frac{1}{7(2m)^7} + \cdots\right] > 0. \tag{8}$$

Therefore, we have

$$\ln\left(\frac{2m+1}{2m-1}\right) = \int_{m-1/2}^{m+1/2}\left(\frac{1}{x}\right)dx > \frac{1}{m}. \tag{9}$$

Summing up for $m = n, n+1, \ldots, N-1$ gives

$$\int_{n-1/2}^{N-1/2}\left(\frac{1}{x}\right)dx > \frac{1}{n} + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{N-1}. \tag{10}$$

For optimal strategies, (5) ensures that the right-hand side of (10) is larger than 1. Thus, the optimal N-value for a fixed S_n strategy is the smallest N that satisfies

$$\ln\left(\frac{2N-1}{2n-1}\right) > 1.$$

Thus, the best N for a given fixed strategy S_n is the least N such that

$$N > e\left(n - \frac{1}{2}\right) + \frac{1}{2} \tag{11}$$

The corresponding probability of winning $P_N(S_n)$ has an upper bound

$$P_N(S_n) \equiv Pr(\Sigma) < Pr(\textstyle\int) \equiv \frac{n}{N}\ln\left(\frac{2N-1}{2n-1}\right). \tag{12}$$

9. THE ASYMPTOTIC BEHAVIOR OF STRATEGIES. Having established two alternative ways of evaluating the probabilities of success for strategies—sums (1) and integrals (12)—we are now in a position to compute these probabilities for large values of N and compare the results. Table 6 displays:

(i) the characteristic optimal (n, N) pairs for large N,
(ii) the respective ration (n/N),
(iii) the corresponding probability of success derived by sums, denoted as $Pr(\Sigma)$, and
(iv) the probability of success derived by the integral approximation, denoted as $Pr(\int)$ in (12).

From Table 6 we see that the probabilities calculated by (1) and (12) both decrease as n increases and they start to (a) coincide (within six significant figures) from the entry ($n = 300$, $N = 815$) onwards, and (b) converge to 0.367879 from the entry ($n =$ 2,000,000, $N =$ 5,436,563). We also observe that the ratio (n/N) converges to the very same number 0.367879. This latter number is e^{-1}.

Thus, we conclude that the two distinct characteristic quantities of optimal strategies, i.e., the ratio (n/N) and the probability of success, both converge to the same limit as $N \to \infty$.

This conclusion may also be derived formally. Since the best N for any S_n strategy is defined as the least N-value satisfying (11), we have $2N > 2en + (1 - e)$ for N, and $2en + (1 - e) \geq 2(N - 1)$ for the immediately lower value, $N - 1$.

TABLE 6. Probabilities of Success for Strategies when N is Large

n	N	n/N	Pr(Σ)	Pr(∫)
100	271	0.369004	0.369045	0.369046
200	543	0.368324	0.368461	0.368462
300	815	0.368098	0.368267	0.368267
400	1087	0.367985	0.368170	0.368170
500	1359	0.367918	0.368112	0.368112
600	1631	0.367872	0.368073	0.368073
700	1902	0.368034	0.368046	0.368046
800	2174	0.367985	0.368025	0.368025
900	2446	0.367948	0.368009	0.368009
1000	2718	0.367918	0.367996	0.367996
2000	5436	0.367918	0.367938	0.367938
3000	8154	0.367918	0.367918	0.367918
4000	10873	0.367884	0.367909	0.367909
5000	13591	0.367891	0.367903	0.367903
6000	16309	0.367895	0.367899	0.367899
7000	19028	0.367879	0.367896	0.367896
8000	21746	0.367884	0.367894	0.367894
9000	24464	0.367888	0.367892	0.367892
10000	27182	0.367891	0.367891	0.367891
20000	54365	0.367884	0.367885	0.367885
30000	81548	0.367881	0.367883	0.367883
40000	108731	0.367880	0.367882	0.367882
50000	135914	0.367880	0.367882	0.367882
60000	163097	0.367879	0.367881	0.367881
70000	190279	0.367881	0.367881	0.367881
80000	217462	0.367880	0.367881	0.367881
90000	244645	0.367880	0.367881	0.367881
100000	271828	0.367880	0.367881	0.367881
200000	543656	0.367880	0.367880	0.367880
300000	815484	0.367880	0.367880	0.367880
400000	1087312	0.367880	0.367880	0.367880
500000	1359141	0.367879	0.367880	0.367880
600000	1630969	0.367879	0.367880	0.367880
700000	1902797	0.367879	0.367880	0.367880
800000	2174625	0.367880	0.367880	0.367880
900000	2446453	0.367880	0.367880	0.367880
1000000	2718281	0.367880	0.367880	0.367880
2000000	5436563	0.367879	0.367879	0.367879

Hence we have

$$\frac{1}{e}\left(1 - \frac{1}{N}\right) \leq \frac{n}{N} - \frac{1}{2eN}\left(\frac{1}{e} - 1\right) < \frac{1}{e}. \tag{13}$$

It follows that $\lim_{N\to\infty} (n/N) = 1/e$ and, consequently, (12) gives $\lim_{N\to\infty} P_N(S_n) = 1/e$.

Concluding, we may state the following *Golden Rule* for decisions:

The optimum strategy is to wait until e^{-1} of the items pass and then select the next relatively best one. The probability of success for the optimum strategy is e^{-1}.

TABLE 7. The Pyramid e-Expansion of N for Optimal Strategies

K	ne − (e − 1) / 2 $\{n = 10^K\}$	n(rounded up)
1	26.32	27
2	270.97	271
3	2717.42	2718
4	27181.96	27182
5	271827.32	271828
6	2718280.97	2718281
7	27182817.43	27182818
8	271828181.99	271828182
9	2718281827.60	2718281828
10	27182818283.73	27182818284
11	271828182845.05	271828182846
12	2718281828458.19	2718281828459
13	27182818284589.59	27182818284590
14	271828182845903.66	271828182845904
15	2718281828459044.38	2718281828459045
16	27182818284590451.49	27182818284590452
17	271828182845904522.68	271828182845904523
18	2718281828459045234.50	2718281828459045235
19	27182818284590452352.74	27182818284590452353
20	271828182845904523535.17	271828182845904523536
21	2718281828459045235359.43	2718281828459045235360
22	27182818284590452353602.02	27182818284590452353603
23	271828182845904523536027.89	271828182845904523536028
24	2718281828459045235360286.61	2718281828459045235360287
25	27182818284590452353602873.85	27182818284590452353602874
26	271828182845904523536028746.28	271828182845904523536028747
27	2718281828459045235360287470.49	2718281828459045235360287471
28	27182818284590452353602874712.67	27182818284590452353602874713
29	271828182845904523536028747134.41	271828182845904523536028747135
30	2718281828459045235360287471351.80	2718281828459045235360287471352
31	27182818284590452353602874713525.77	27182818284590452353602874713526
32	271828182845904523536028747135265.39	271828182845904523536028747135266
33	2718281828459045235360287471352661.64	2718281828459045235360287471352662
34	27182818284590452353602874713526624.12	27182818284590452353602874713526625
35	271828182845904523536028747135266248.92	271828182845904523536028747135266249
36	2718281828459045235360287471352662496.90	2718281828459045235360287471352662497
37	27182818284590452353602874713526624976.71	27182818284590452353602874713526624977
38	271828182845904523536028747135266249774.87	271828182845904523536028747135266249775
39	2718281828459045235360287471352662497756.39	2718281828459045235360287471352662497757
40	27182818284590452353602874713526624977571.61	27182818284590452353602874713526624977572

Having established that the golden rule for decisions is associated with e, it is interesting to note that when n takes as values the successive integer powers of 10, the corresponding values of N generate the decimal expansion of e. This is illustrated in Table 7.

ACKNOWLEDGMENTS. We thank our students Miss A. Demosthenous, Mr. A. Giouzenis, Miss E. Hondrou, and Mr. C. Nakos for their contributions to the editing of this work.

REFERENCES

1. B. H. Bissinger and C. Siegal, Problem 5086, *Amer. Math. Monthly* **70** (1963) 336.
2. A. J. Bosch, Solution to Problem 5086: Optimum strategy in a guessing game, *Amer. Math. Monthly* **71** (1964) 329–330.

3. Y. S. Chow, S. Moriguti, H. Robbins, and S. M. Samuels, Optimal Selection based on relative rank (The "Secretary" problem), *Israel J. Math* **2** (1964) 81–90.
4. E. B. Dynkin, The optimum choice of the instant for stopping a Markov process, *Soviet Math. Dokl.* **4** (1963) 627–629.
5. T. S. Ferguson, Who solved the secretary problem?, *Statist. Sci.* **4** (1989) 282–296.
6. P. R. Freeman, The secretary problem and its extensions—A review, *Internat. Statist. Rev.* **51** (1983) 189–206.
7. M. Gardner, *Scientific American* **202** (1960) 150–153.
8. J. Gilbert and F. Mosteller, Recognizing the maximum of a sequence, *J. Amer. Statist. Assoc.* **61** (1966) 35–73.
9. D. V. Lindley, Dynamic programming and decision theory, *Appl. Statist.* **10** (1961) 39–51.

DIMITRIS A. SARDELIS received his Ph.D from Imperial College where he worked on quantum field theories under the supervision of Abdus Salam. He has been teaching at the American College of Greece since 1991. He is interested in the foundations of geometry, stochastic processes, and higher arithmetic.

THEODOROS M. VALAHAS received his B.S. from Aristoteles University of Thessaloniki, Greece and his Ph.D. from Massachusetts Institute of Technology. After spending some years in industry he returned to the academic community. He co-ordinates the math and science group of the American College of Greece. His interests are classical fundamental problems in kinematics, dynamics, number theory, probabilities, and statistical decision theory.
Department of Sciences & Mathematics, The American College of Greece, Aghia Paraskevi, Athens GR-15342, Greece
dereef@hol.gr

There is a sibling rivalry
between this conjecture and its negation
and I, poor mother
throw up my hands.
"Anything, anything.
"Whatever you decide.
"Just please
"hurry up
"and make up your mind."

Contributed by Marion Cohen, Drexel University, Philadelphia, PA

A Tale of Two Integrals

Vilmos Totik

1 THE PROBLEM. We present several approaches to a simple-looking but highly nontrivial combinatorial–analysis problem. Our aim is to show how different ideas can lead to a solution.

The problem is easy to state: *Let f and g be two integrable functions on* $[0, 1]$ *with*

$$\int_0^1 f = \int_0^1 g = 1. \tag{1}$$

Show that there is some interval $I \subset [0, 1]$ *such that*

$$\int_I f = \int_I g = \frac{1}{2}. \tag{2}$$

Instead of $[0, 1]$ we could have any interval, and f and g need not have integrals equal to 1; the general statement is that there is always a single interval where each function has integral equal to one half of its total integral.

Here is an equivalent formulation without integrals: *On a blackjack machine one can win or lose one dollar at a time. Suppose two players playing once per minute during a period find that eventually both of them win exactly* $2N$ *dollars. Show that there was a time interval during which both of them won exactly* N *dollars.*

We sketch the equivalence of the two forms. The second form is a consequence of the first if we apply it to some appropriate step functions f and g with values ± 1 modelling the outcome of the blackjack games (see Figure 1). Then there is an interval (a, b) over which both f and g have integral N. If both a and b are integers, then going back to the blackjack game we get a time interval with the desired properties. If they are not, then with some $0 < \alpha < 1$ we have $a = [a] + \alpha$

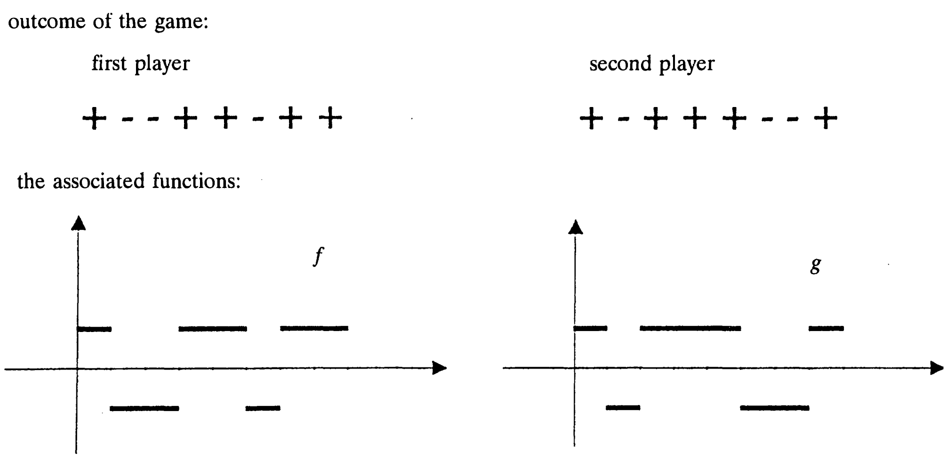


Figure 1. The equivalence of the two formulations

and $b = [b] + \alpha$. If $\alpha \neq \frac{1}{2}$, then the integrals of f and g over $([a],[b])$ are again equal to N, so we are back at the integer case. This is also true if $\alpha = \frac{1}{2}$ and both functions have the same sign at a and b. Finally, if $\alpha = \frac{1}{2}$ and one of them has different signs at a and b, then a simple parity argument (based on the fact that two sums of the form $\sum_{k=1}^{m} \pm 1$ are equal or differ by an even integer) shows that the same is true of the other function, and then the integrals over $([a]+1,[b])$ are again N.

In the opposite direction suppose the blackjack statement is true, and let f and g be two functions satisfying (1). We can assume (see below) f and g to be bounded, say $|f| \leq M$, $|g| \leq M$. Then the graphs of the functions

$$\frac{1}{M}\int_0^x f \qquad \text{and} \qquad \frac{1}{M}\int_0^x g$$

can be arbitrarily well approximated by piecewise linear curves with equidistant nodes and slopes ± 1. Now the slope functions of these curves can be regarded as the outcome of blackjack games for two players ($+1$ stands for winning and -1 for losing a dollar), so the second formulation can be applied. Going from here to (2) is a routine limiting process to be discussed below.

We shall adhere to the first formulation, though the precise notion of "integrable" is irrelevant. In fact, the problem is not easier if we assume that f and g are continuous, or that they are step functions. To see this it is enough to note the following. Suppose that f_n and g_n are functions satisfying (1) such that

$$\int_0^1 |f_n - f| \to 0 \qquad \text{and} \qquad \int_0^1 |g_n - g| \to 0,$$

and suppose we can verify the existence of intervals $I_n = (a_n, b_n)$ satisfying (2) for the pairs f_n, g_n. By selecting a subsequence $\mathbf{N}_1$ of the natural numbers for which $\{a_n\}_{n \in \mathbf{N}_1}$ and $\{b_n\}_{n \in \mathbf{N}_1}$ converge to some a and b, we can show that (2) holds for $I = (a, b)$. Thus, without loss of generality we may assume that f and g belong to any chosen dense subspace of the space of integrable functions, for example the space of continuous functions or the space of step functions.

We present several solutions to the problem that are related to other combinatorial or geometrical/topological results. Some of these solutions are genuinely different, some are interrelated, but all of them use some well known facts of mathematics. Other approaches are also known, but none of the elementary solutions (that mainly use induction) I know is short enough to present in one or two pages.

The problem appeared on the 1995 Miklós Schweitzer Mathematical Contest in Hungary. This is a unique mathematical contest organized every fall since 1949 by the János Bolyai Mathematical Society. It is a contest for university students and fresh graduates, but sometimes talented high school students also successfully participate. About a dozen problems (almost exclusively new) are proposed from different branches of mathematics, and the students have 10 days to solve them using all available literature. Accordingly, the problems are considerably more difficult than on other mathematical competitions and olympiads. The problems and solutions from the contests in the years 1962–1991 have been published in the Springer Problem Book series: *Contests in Higher Mathematics*, Ed. G. J. Székely, 1995.

The problem is more difficult than one might expect. In fact, the very first thought, namely continuously move the endpoints a and b of $I = (a, b)$, leads to

big obstacles. The argument would run like this: there is an a_0 such that for each $0 \le a \le a_0$ there is a b_a with $\int_a^{b_a} f = \frac{1}{2}$. If $\int_0^{b_0} g = \frac{1}{2}$, then we are done. If, say, $\int_0^{b_0} g < \frac{1}{2}$, then we must have $\int_{b_0}^1 g > \frac{1}{2}$. Therefore if we continuously move a from 0 to b_0, there will be a value of a for which the integral $\int_a^{b_a} g$ is exactly $\frac{1}{2}$, and so $I = (a, b_a)$ is suitable. The problem with this reasoning is that b_a does not depend continously on a, and the whole argument collapses. Even worse than that, in general there is no continuous function $b(a)$ such that

$$\int_a^{b(a)} f = \frac{1}{2} \tag{3}$$

for all $a \in [0, a_0]$ (see Figure 2). Thus, the preceding reasoning cannot be rectified.

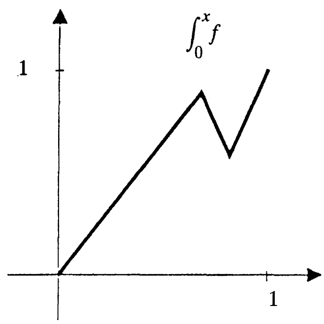


Figure 2. No continuous $b(a)$ exists

This simple continuity argument does work if f is strictly positive, for then there is a single $b(a)$ satisfying (3), and the function $b(a)$ is continuous. By adding ϵ to f and then letting $\epsilon \to 0$ we can conclude that the same is true for nonnegative f. Let us also mention that for piecewise constant f *there are continuous functions* $a(t)$, $b(t)$ on the parameter interval $[0, 1]$ such that

$$\int_{a(t)}^{b(t)} f = \frac{1}{2}, \tag{4}$$

$a(0) = 0$, $a(1) = b(0)$, and $b(1) = 1$, so a continuity argument like the one before can be applied. However, proving the existence of $a(t)$ and $b(t)$ is as difficult as the original problem.

2 THE BORSUK–ULAM ANTIPODAL THEOREM. The Borsuk–Ulam Theorem [**1**, p. 241] states that *if* $T: S^2 \to \mathbf{R}^2$ *is a continuous mapping of* S^2 *(the unit sphere in* R^3*) into the plane, then there exists a pair of antipodal points* $\{X, -X\}$ *of* S^2 *that have the same image*: $T(X) = T(-X)$. If T is also odd, i.e., $T(-Y) = -T(Y)$ for all $Y \in S^2$, then we must have $T(X) = (0, 0)$. The same is true in higher dimensions, namely if $T: S^l \to \mathbf{R}^l$ is a continuous mapping of the unit sphere in $\mathbf{R}^{l+1}$, then there exists a pair of antipodal points $\{X, -X\}$ on S^l that have the same image.

Using the Borsuk–Ulam Theorem, the solution of our problem is easy. Let (ξ_1, ξ_2, ξ_3), $\xi_1^2 + \xi_2^2 + \xi_3^2 = 1$ be a point on S^2, and let

$$T(\xi_1, \xi_2, \xi_3) = (X(f; \xi_1, \xi_2, \xi_3), X(g; \xi_1, \xi_2, \xi_3)),$$

where

$$X(f; \xi_1, \xi_2, \xi_3) = \operatorname{sign}(\xi_1) \int_0^{\xi_1^2} f + \operatorname{sign}(\xi_2) \int_{\xi_1^2}^{\xi_1^2 + \xi_2^2} f + \operatorname{sign}(\xi_3) \int_{\xi_1^2 + \xi_2^2}^{1} f. \tag{5}$$

Since T is a continuous odd mapping of S^2 into the plane, the Borsuk–Ulam theorem ensures that some point $(\xi_1^*, \xi_2^*, \xi_3^*)$ is mapped into $(0,0)$ by T. Among the numbers $\xi_1^*, \xi_2^*, \xi_3^*$, two have the same sign (consider 0 to be of positive sign). If the third number is ξ_j^*, and I denotes the interval of length ξ_j^{*2} in the integral multiplied by sign (ξ_j^*) in (5), then from the definition of T and from $T(\xi_1^*, \xi_2^*, \xi_3^*) = (0,0)$ it follows that

$$\int_I f = \int_{[0,1]\setminus I} f \quad \text{and} \quad \int_I g = \int_{[0,1]\setminus I} g.$$

The claim follows with this I if we also take into account the conditions $\int_0^1 f = \int_0^1 g = 1$.

The Borsuk–Ulam theorem is a standard tool in solving the following necklace of pearls problem. *Two pirates have a single-strand necklace containing $2k$ black pearls and $2k$ white pearls arranged in any order. They would like to cut the necklace into as few pieces as possible so that after dividing the pieces of the necklace between them, each gets exactly k white pearls and k black ones.* An easy modification of the preceding solution permits us to conclude that two cuts are always enough (i.e., there is always a sequence of $2k$ consecutive pearls on the necklace that contains k pearls of each type). This is related to the case of our original problem where both functions (representing the two types of pearls on intervals of equal lengths, see Figure 3) are nonnegative; hence we do not need the antipodal theorem, as

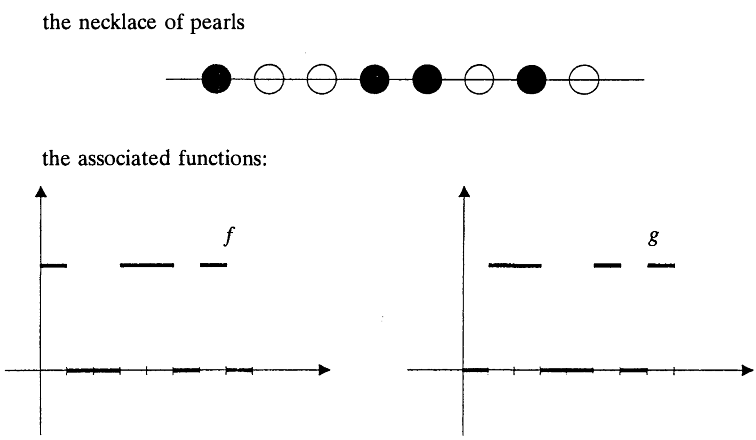


Figure 3. Reduction of the pearl problem

the solution follows by a simple continuity argument. If the necklace contains l types of pearls, and there are $2k$ pearls of each type, we can apply the higher dimensional version of the Borsuk–Ulam theorem to show that l cuts are always enough.

In a similar manner, by applying the antipodal theorem in higher dimensions we get the following generalization of our original problem due to A. Pinkus: *if $f_1, \dots, f_l \in L^1[0,1]$ and $\int_0^1 f_j = 1$ for all $j = 1, \dots, l$, then there is a set I consisting of at most $(l+1)/2$ intervals such that $\int_I f_j = \frac{1}{2}$ for all $j = 1, \dots, l$.*

3 THE MOUNTAIN CLIMBING PROBLEM. Can two climbers climb up opposite sides of a mountain to the top in such a way that both of them are always at the same altitude (see [7] and [9])? There are some obvious obstacles that prevent

them from doing so, however, the answer is YES if the sides of the mountain are piecewise linear curves and the climbers start at the bottom **[6]**. We show that this result implies a solution to our problem.

Assume, as we may, that both f and g are piecewise constant. Then

$$H(x) = \int_0^x (f(u) - g(u))\, du$$

is piecewise linear, and $H(0) = H(1)$, so we can extend H to a continuous 1-periodic function. Let us also extend f and g periodically to $\mathbf{R}$ with period 1. The graph of H is our "mountain" (see Figure 4), and two climbers climb from the

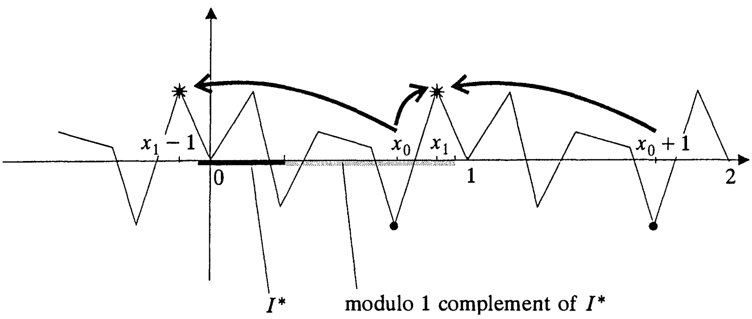


Figure 4. The mountain of the two climbers

bottom level of that mountain, say from the points x_0 and $x_0 + 1$, to one of the maximum points of H in $[x_0 . x_0 + 1]$, say to the peak at x_1. According to **[6]** they can do so and stay always at the same level. Now by the periodicity, we can assume that both climbers start at x_0, the first one climbs to the right to the peak at x_1, while the second one climbs to the left to the peak at $x_1 - 1$. Let the x coordinate of the two climbers at time $t \in [0, 1]$ be $\gamma_1(t)$ and $\gamma_2(t)$. Thus, the γ_j are continuous functions such that $\gamma_1(0) = \gamma_2(0) = x_0$, $\gamma_1(1) = x_1$, $\gamma_2(1) = x_1 - 1$, and at every moment $\gamma_2(t) \le x_0 \le \gamma_1(t)$. Since the climbers always stay at the same altitude, we have

$$\int_{\gamma_2(t)}^{\gamma_1(t)} f(u)\, du = \int_{\gamma_2(t)}^{\gamma_1(t)} g(u)\, du. \tag{6}$$

However, the left integral is zero for $t = 0$, is 1 for $t = 1$ since $[\gamma_2(1); \gamma_1(1)] = [x_1 - 1, x_1]$ is a full period for f, and is a continous function of t, so there is a $t = t^*$ for which the left-hand side is equal to $\frac{1}{2}$. But then the right hand side is also $\frac{1}{2}$, which means that for $I^* = [\gamma_2(t^*), \gamma_1(t^*)]$ we have

$$\int_{I^*} f = \int_{I^*} g = \frac{1}{2}.$$

This seems to be what we are looking for, but we have to be careful, for the interval I^* may not belong to $[0, 1]$ (or for that matter to some $[n, n + 1]$ with integer n). If it does, then we just set $I = I^*$. If not, then we can take as I its complement in $[0, 1]$ modulo 1 (see Figure 4), which, in view of (1), satisfies (2).

This solution yields the following generalization. *Let f and g be integrable functions on* $[0, 1]$ *satisfying* (1), *and let* $0 < \alpha < 1$. *If there is no interval* $I \subset [0, 1]$

with

$$\int_I f = \int_I g = \alpha,$$

then there is an interval I with

$$\int_I f = \int_I g = 1 - \alpha. \tag{7}$$

Our proof gives $I = I^*$ for α if $I^* \subset [0,1]$; if $I^* \not\subset [0,1]$, then its modulo 1 complement is suitable for I in (7).

Thus, for any given $\alpha \in (0,1)$, there is always an interval I where the integrals of both f and g equal either α or $1 - \alpha$. For $\alpha = \frac{1}{2}$ this means that both f and g have integrals $\frac{1}{2}$, as is claimed in the problem. For $\alpha = \frac{1}{3}$ there is an interval I such that either both functions have integral $\frac{1}{3}$ on I, or both of them have integral $\frac{2}{3}$. In the latter case we can apply the already proven $\frac{1}{2}$-case to I, and conclude that on some subinterval of I both functions have integral equal to $(2/3)/2 = \frac{1}{3}$. Thus, there is an interval with both integrals equal to $\frac{1}{3}$. This argument can be repeated, to obtain an interval over which both integrals equal any given value $\frac{1}{4}, \frac{1}{5}, \ldots$. This is a generalization of the original problem: *Let f and g be two integrable functions on* $[0,1]$ *satisfying* (1) *and let k be a positive integer. Then there is an interval I such that*

$$\int_I f = \int_I g = \frac{1}{k} \tag{8}$$

This is false for every $\alpha \in (0,1)$ that is not of the form $\alpha = 1/k$. In fact, if $f(x) = (2n+1)/(n+1)$ if $2k/(2n+1) \le x \le (2k+1)/(2n+1)$ $(k = 0, 1, \ldots, n)$ and 0 otherwise, and if $g(x) = (2n+1)/n$ if $(2k-1)/(2n+1) \le x \le 2k/(2n+1)$ $(k = 1, 2, \ldots, n)$ and 0 otherwise, then for no $\alpha \in ((n+1)^{-1}, n^{-1})$ is there an interval I for which

$$\int_I f = \int_I g = \alpha$$

(see Figure 5 where the $n = 1$ case is displayed).

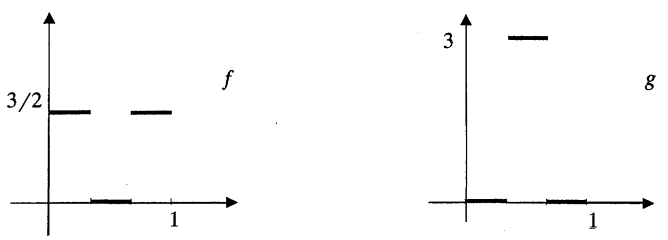


Figure 5. No common intervals for $\frac{1}{2} < \alpha < 1$

4 THE CHORD THEOREM. The chord theorem [3, pp. 21 and 198–199] states that *if γ is a continuous curve on the plane with endpoints A and B, then for every positive integer k there is a chord $\overline{CD}$ of γ (i.e., $C, D \in \gamma$) that is parallel to $\overline{AB}$ and has length $1/k$ times the length of $\overline{AB}$.*

Apply the chord theorem to the curve

$$\gamma(t) := \left(\int_0^t f(u)\, du, \int_0^t g(u)\, du \right), \qquad t \in [0,1]$$

with $k = 2$. Since γ has endpoints $(0,0)$ and $(1,1)$, it has a chord of the form $(X,Y)(X+\frac{1}{2}, Y+\frac{1}{2})$, i.e., if we choose parameters $t_1, t_2 \in [0,1]$ such that $\gamma(t_1) = (X,Y)$ and $\gamma(t_2) = (X+\frac{1}{2}, Y+\frac{1}{2})$, then

$$\int_{t_1}^{t_2} f(u)\, du = \int_{t_1}^{t_2} g(u)\, du = \frac{1}{2}, \tag{9}$$

and we seem to have solved the problem. However, I. Z. Ruzsa observed that we might have $t_2 < t_1$, i.e., if I is the interval determined by the parameters t_1 and t_2 (which is $[t_2, t_1]$ for $t_2 < t_1$), then (9) means that on I both functions have integral $-\frac{1}{2}$, instead of $\frac{1}{2}$, so the chord theorem does not give us what we want.

Our approach via the chord theorem can be saved as follows. Assume, as we may, that f and g are piecewise constant functions that do not vanish on any subinterval. Select a maximal subinterval $J_1 \subset [0,1]$ such that the integrals of f and g over J_1 are equal, and the common value of the integrals is either zero or a positive integer multiple of $-\frac{1}{2}$. Then select a maximal subinterval $J_2 \subset [0,1]$ *disjoint from* J_1 such that the integrals of f and g over J_2 are equal, and the common value of these integrals is either zero or a positive integer multiple of $-\frac{1}{2}$. Continue this process by always selecting maximal subintervals that are disjoint from all previously selected subintervals. We claim that this process must terminate in finitely many steps. If not, then there was an infinite family of disjoint subintervals over which the integral of f is zero or a positive integer multiple of $-\frac{1}{2}$. That the integral cannot be a positive multiple of $-\frac{1}{2}$ for infinitely many disjoint subintervals is clear. Hence, there are infinitely many disjoint intervals over which f has zero integrals. However, this is again impossible, because every such interval must contain a discontinuity point of f; recall that f is piecewise constant with nonzero function values.

Let the selected maximal intervals be $J_1, \dots, J_m$. Contract each of the J_k's to a single point (see Figure 6). We obtain an interval $[0, a]$, an integer $k \geq 2$, and two functions f^* and g^* such that

$$\int_0^a f^* = \int_0^a g^* = \frac{k}{2}.$$

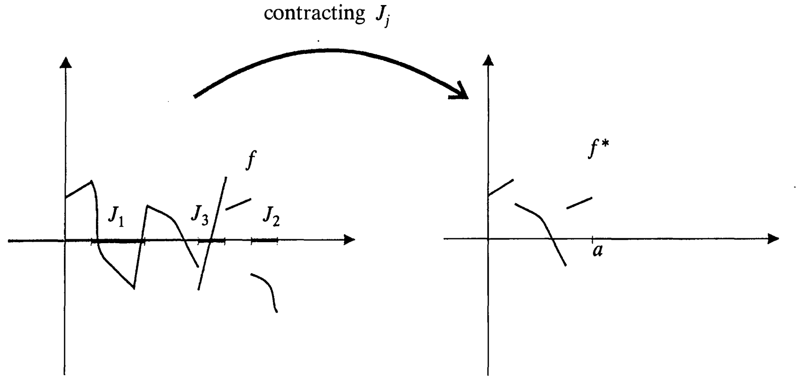


Figure 6. Contracting the J_k's

Note that we have removed subintervals from $[0,1]$ over which the integrals were a nonnegative integer multiple of $-\frac{1}{2}$. Now apply the $1/k$ version of the chord theorem to this pair. As before, we get a subinterval I^* of $[a,b]$ such that f^* and g^* have equal integrals over I^*, and the common value of these integrals is $\pm\frac{1}{2}$. Let I be the subinterval of $[0,1]$ that corresponds to I^* under the contraction when we removed the intervals J_k. If the integral over I^* of f^* and g^* is $-\frac{1}{2}$, then I cannot contain J_1 by the maximality of J_1, so I and J_1 are disjoint. For the same reason I cannot contain J_2, or J_3, etc. Thus, I is a subinterval disjoint from every J_k over which both f and g have integral $-\frac{1}{2}$, which is impossible, since the system $\{J_k\}_{k=1}^m$ is maximal.

Therefore, the integral of f^* and g^* over I^* is $\frac{1}{2}$. But then I can contain only J_k's over which the integral of f and g is zero, for otherwise the integral over I would be zero or a positive multiple of $-\frac{1}{2}$, and this would contradict the maximality of the first J_k that is contained in I. Therefore, I contains only intervals J_k over which both f and g have integrals zero, so the integrals of f and g over I are the same as those of f^* and g^* over I^*, which is $\frac{1}{2}$. Therefore, this I is suitable.

5 THE CHESS KING–MOVING THEOREM. Suppose we color the squares of an $n \times n$ (chess) board with black and white arbitrarily. The chess king moving theorem (see **[4]**, **[5]**) asserts that *a chess king can move either from the top row to the bottom row on black squares, or it can move from the leftmost column to the rightmost column on white squares*.

This statement is strong enough to prove the Brouwer fixed point theorem in two dimensions **[5]**. It also shows that in the following very entertaining game there is always a winner: two players B and W place alternately black and white disks on an $n \times x$ board. B's aim is to connect the upper and lower edges of the board by his black disks, while W wants to connect the left and right sides of the board by his white disks. The game Hex is identical to this one, except that it is played on a rhomboid-shaped board of hexagons.

Now let us see how the chess king–moving theorem solves our problem. We again extend the functions f and g to the whole real line as 1-periodic functions, and, as we have already seen, it is sufficient to verify the existence of an interval I of length less than 1 somewhere on $\mathbf{R}$ satisfying property (2). The function

$$\int_0^x f - \int_0^x g$$

attains its minimum at some point a. This implies that

$$\int_a^x f - \int_a^x g \geq 0$$

for all $x \geq a$. Therefore, by replacing f and g by $f^*(y) = f(a+y)$ and $g^*(y) = g(a+y)$, we can assume that

$$\int_0^y f - \int_0^y g \geq 0 \tag{10}$$

for all $y \geq 0$. Since the integral of both functions over an interval of length 1 is 1, this implies that for all $y \in [0,1]$

$$\int_y^1 f - \int_y^1 g \leq 0. \tag{11}$$

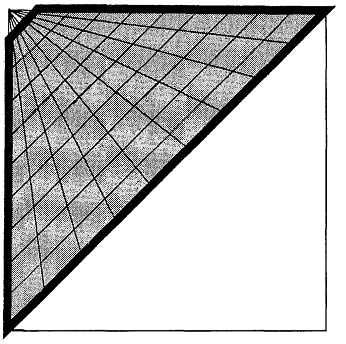


Figure 7. The squeezed chess board

Fix $\epsilon > 0$. Partition the triangle lying above the diagonal in the unit square by $n-1$ rays emanating from the point $(0,1)$ and by n lines parallel to the diagonal as illustrated in Figure 7. Discard the small triangles that contain the point $(0,1)$. The remaining figure, partitioned into the (closed) pieces $J_1, \ldots, J_{n^2}$, will be our chess board; it is a squeezed chess board, but the chess king-moving theorem is insensitive to the actual shape of the cells on the board. If n is sufficiently large, then for points (x_i, y_i) and (x_j, y_j) lying in neighboring cells J_i and J_j we have

$$\left|\int_{x_i}^{y_i} f - \int_{x_j}^{y_j} f\right| \leq \epsilon \qquad \text{and} \qquad \left|\int_{x_i}^{y_i} g - \int_{x_j}^{y_j} g\right| \leq \epsilon. \tag{12}$$

Now color a cell J black if there is a point $(x, y) \in J$ with

$$\left|\int_x^y f - \int_x^y g\right| \leq \epsilon; \tag{13}$$

all other cells remain white. If a cell J is white and is in the leftmost column, then by (10) we must have

$$\int_x^y f - \int_x^y g > \epsilon$$

for all $(x, y) \in J$, and conversely, (11) shows that if a white cell J is in the rightmost column (which means that it is on the upper edge of the unit square), then

$$\int_x^y f - \int_x^y g < -\epsilon$$

for all $(x, y) \in J$. Furthermore, in neighboring white cells the difference

$$\int_x^y f - \int_x^y g$$

can change by a quantity with absolute value at most 2ϵ (see (12)). Thus, by the definition of the coloring, a king cannot move on white cells from the leftmost column to the rightmost one. Therefore, it can move from the upper row to the lower one on black cells.

However, the cells in the upper row are around the point $(0,1)$, so for points (x, y) in those cells the integral

$$\int_x^y f \tag{14}$$

is close to $\int_0^1 f = 1$, while cells in the bottom row contain diagonal points $(x, y) = (x, x)$ for which (14) is zero. Therefore, by (12), as the king moves from the upper row to the bottom row, it must pass through a black cell whose points satisfy

$$\left| \int_x^y f - \frac{1}{2} \right| \leq 2\epsilon,$$

and since this is a black cell, we also have

$$\left| \int_x^y g - \frac{1}{2} \right| \leq 5\epsilon.$$

What we have found is that for every $\epsilon = 1/n$, $n = 1, 2, \dots$ there are points $0 \leq x_n \leq y_n \leq 1$ such that

$$\left| \int_{x_n}^{y_n} f - \frac{1}{2} \right| \leq \frac{2}{n} \quad \text{and} \quad \left| \int_{x_n}^{y_n} g - \frac{1}{2} \right| \leq \frac{5}{n}.$$

From here the rest is standard: select a subsequence $\mathbf{N}_1$ of the natural numbers for which $\{x_n\}_{n \in \mathbf{N}_1}$ converges to some x and at the same time $\{y_n\}_{n \in \mathbf{N}_1}$ converges to some y. Then the preceding inequalities easily yield

$$\int_x^y f = \int_x^y g = \frac{1}{2},$$

which is what we had to verify.

6 THE WINDING NUMBER THEOREM. *Let $\Delta \subset \mathbf{R}^2$ be the unit disk and let $V: \Delta \to \mathbf{R}^2$ be a continuous vector field that does not vanish on the circumference. If the winding number of V on the circumference is not zero, then V must vanish somewhere on Δ* [**2**, pp. 134–135] or [**1**, pp. 255–257].

The term "vector field" comes from the fact that for every point P of Δ the value $V(P)$ is a two dimensional vector, and we can think of placing the tail of this vector at the point P (see Figure 8). As we move around the circumference, the

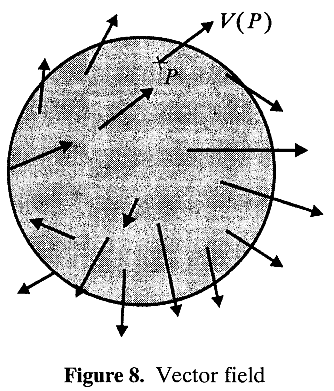


Figure 8. Vector field

vector $V(\cos t, \sin t)$ is not zero and depends continously on $t \in [0, 2\pi]$, so the angle it forms with the positive half axis is also a continuous function $A(t)$; we *do not take* the angle modulo 2π. For the parameter value $t = 2\pi$ we arrive back again at the point $(0, 1)$ associated with $t = 0$, so the vectors corresponding to these two parameter values are the same. Thus, $A(2\pi)$ must be equal to $A(0)$ plus a positive multiple of 2π. The *winding number* of V is $(A(2\pi) - A(0))/2\pi$, which is an integer.

The winding number theorem can be proved as follows: If the field does not vanish on Δ, then the winding number on circles $\{|z| = a\}$ changes continuously with a. Since it is always an integer, it is constant. However, this constant is not zero for $a = 1$ by assumption, and it is clearly zero for $a = 0$. This contradiction shows that the vector field must vanish somewhere.

Let us see how this theorem solves the problem (a solution by Attila Pór). Consider the subtriangle $D = \{(x, y)|0 \leq x \leq y \leq 1\}$ lying above the diagonal of the unit square, and for $(x, y) \in D$ let

$$U(x, y) = \left(\int_x^y f - \frac{1}{2}, \int_x^y g - \frac{1}{2}\right).$$

We have to show that the vector field U vanishes somewhere in D.

Let $\varphi : \Delta \to D$ be a continuous one-to-one mapping between the unit disk and D (a homeomorphism). Then $V = U \circ \varphi$ defines a vector field on Δ, and the vanishing of U is equivalent to the vanishing of V. If V vanishes somewhere on the boundary, then we are done. If not, then V defines a continuous vector field on Δ that does not vanish on the boundary. If its winding number is not zero, the winding number theorem ensures that V vanishes somewhere on Δ. Since φ carries the vectors from the field U into the vectors of the field V, we can work directly on D, where the winding number of the field U is defined as the winding number of V.

On the diagonal of the triangle the vector field U has the constant value $(-\frac{1}{2}, -\frac{1}{2})$. Therefore, on this part of the boundary the field U does not rotate. If $0 \leq x \leq 1$, then (1) shows that

$$U(0, x) + U(x, 1) = (0, 0),$$

i.e.,

$$U(x, 1) = -U(0, x).$$

It follows that the total winding of the field along the horizontal side of D is the same as the total winding along the vertical side (both travelled, say, in counter-clockwise direction), because the corresponding angles always differ by π. Furthermore, since $U(0, 0) = (-\frac{1}{2}, -\frac{1}{2})$, while $U(0, 1) = (\frac{1}{2}, \frac{1}{2})$, the angles of these two vectors must differ by $2k\pi + \pi$ for some integer k. Therefore, the winding number of the field U along the boundary of D is

$$2(2k\pi + \pi)/2\pi = 2k + 1 \neq 0,$$

and this is what we needed to prove.

7 THE JORDAN CURVE THEOREM. The Jordan curve theorem [**8**] states that *any continuous simple closed curve on the plane divides the plane into two connected components*.

To our problem we give a solution, due to Tamás Fleiner, that relies on an intuitively simple fact. We need the Jordan curve theorem to verify formally the intuitively obvious part.

Assume, as we may, that both f and g are step functions. Extend both f and g to $\mathbf{R}$ as periodic functions with period 1, and let

$$F(x) = \int_0^x f \quad \text{and} \quad G(x) = \int_0^x g.$$

Then for all x we have $F(x + 1) = F(x) + 1$ and $G(x + 1) = G(x) + 1$. Furthermore, F and G are continuous, so for the points $a, b \in [0, 1]$ where $F - G$ attains

its maximum and minimum, respectively, we have for all x

$$A := F(a) - G(a) \geq F(x) - G(x) \qquad \text{and}$$
$$B := F(b) - G(b) \leq F(x) - G(x).$$

The curve $\gamma(x) = (F(x), G(x))$ is a union of line segments. We verify that there is a $y \in [0,1]$ and an $y - 1 < x < y$ such that $\gamma(y) = \gamma(x) + (\frac{1}{2}, \frac{1}{2})$. If $A = B$, then $F \equiv G$ and we have the trivial case $f \equiv g$. If $A \neq B$, then by interchanging the role of F and G, respectively, we may assume $a < b$. The curve $\gamma(x)$ lies within the strip S determined by the lines $x - y = A$ and $x - y = B$, and the portion of γ corresponding to the parameter values $x \in [a, b]$ connects the two bounding lines of this strip (see Figure 9). By replacing a by the largest value

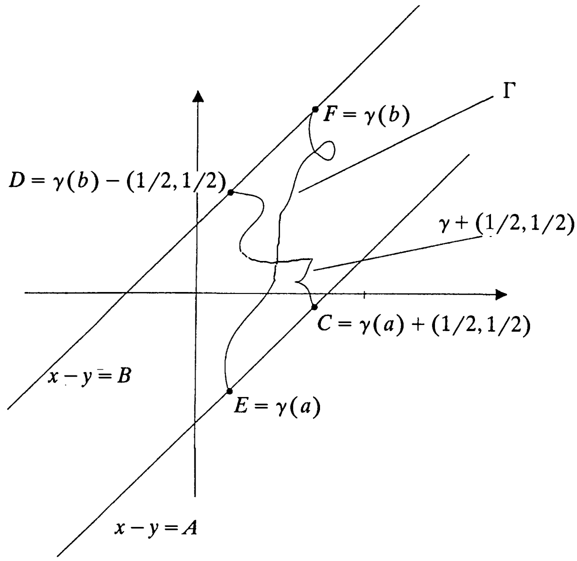


Figure 9. The two paths CD and EF have to intersect

$a' < b$ for which $\gamma(a')$ belongs to the lower bounding line $x - y = A$, and then replacing b by the smallest $a' < b'$ for which $\gamma(b')$ belongs to the upper bounding line $x - y = B$, we can also suppose that the curve

$$\Gamma := \{\gamma(x) | a \leq x \leq b\}$$

lies strictly within the strip S except for its two endpoints. Now the points $C := \gamma(a) + (\frac{1}{2}, \frac{1}{2})$ and $D := \gamma(b) - (\frac{1}{2}, \frac{1}{2})$ cannot be connected by a continuous piecewise-linear path that does not leave S and does not intersect the curve Γ (look at the X-shaped figure in Figure 9). However, $\gamma(b) - (\frac{1}{2}, \frac{1}{2})$ is $\gamma(b - 1) + (\frac{1}{2}, \frac{1}{2})$, so the two points C and D *do lie* on the curve $\gamma(x) + (\frac{1}{2}, \frac{1}{2})$, $b - 1 < x < a$. Hence, there must be a point of intersection, i.e., there is an $x \in (b - 1, a)$ and a $y \in (a, b)$ such that $\gamma(y) = \gamma(x) + (\frac{1}{2}, \frac{1}{2})$. Furthermore, $b - 1 < x < a < y < b$, so $y - 1 < x < y$. By the definition of the curve γ, $F(y) - F(x) = \frac{1}{2}$ and $G(y) - G(x) = \frac{1}{2}$, i.e., both f and g have integral $\frac{1}{2}$ on the interval $[x, y]$. The rest of the argument is now the same as in Section 4: if $x \geq 0$, then the interval $I = [x, y]$ satisfies the requirements. If, however, $x < 0$, then the interval $I = [y, x + 1]$ is

suitable, for then

$$\int_I f = F(x+1) - F(y) = 1 + F(x) - F(y) = 1 - \frac{1}{2} = \frac{1}{2},$$

and a similar calculation shows that the integral of g over I is again $\frac{1}{2}$.

This proof is based on the fact that two curves in S, one connecting the points E and F and the other connecting the points C and D, must intersect. This is intuitively clear, but for a formal verification we invoke the Jordan curve theorem stating that any continuous simple closed curve τ on the plane divides the plane into two connected components. By a continuous simple closed curve we mean a continuous function $\tau:[0,1] \to \mathbf{R}^2$ such that $\tau(0) = \tau(1)$, and for $0 \le x < y < 1$ we have $\tau(x) \ne \tau(y)$ (i.e., the points on the curve are all different, except for the starting and ending points). The statement itself means that $\mathbf{C} \setminus \tau = U \cup V$, where every point of U can be connected to any other point of U by a continuous piecewise-linear path lying in U, and similarly for V. Furthermore, no two points lying in U and V, respectively, can be connected by such a path not intersecting τ. How do we know that we cross from one component (U or V) to another one? Well, this is certainly the case if we move along a segment that intersects τ in exactly one point, and is perpendicular to a segment of τ.

The fact that Γ has a non-empty intersection with any continuous piecewise-linear path connecting C and D within S now can be verified as follows. By removing loops from Γ we can assume that it is a simple curve. Then consider the curve τ described in Figure 10. Moving along the segment CH we get from one connected

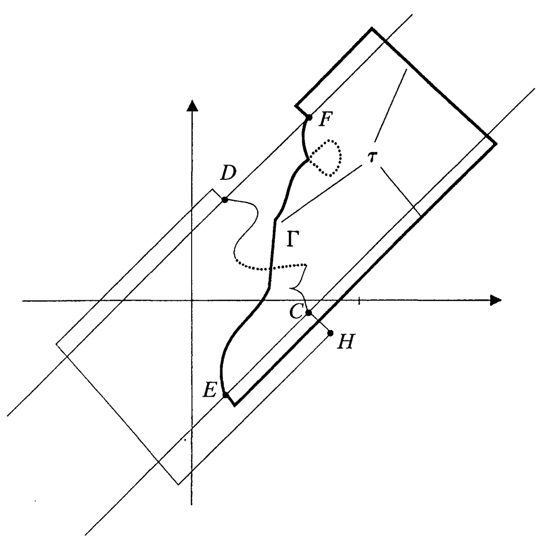


Figure 10. The two paths CD and EF must intersect

component of $\mathbf{C} \setminus \tau$ to the other one, so the points C and D are in different components of $\mathbf{C} \setminus \tau$. Therefore, any continuous piecewise-linear path connecting C and D lying in S (e.g., $\{\gamma(x) | b - 1 < x < a\}$) must intersect τ. But $\tau \cap S = \Gamma$, so every such path must intersect Γ itself, which is what we needed to prove.

We have presented several solutions to our problem that were based on some known theorems from planar geometry and topology. Some of these theorems are

also interrelated and it is easy to see that the statement in our problem is actually equivalent to at least one of them, namely to the chord theorem.

ACKNOWLEDGMENT. The author acknowledges partial support by the National Science Foundation (DMS–9415657) and by the Hungarian Academy of Sciences (96–328).

REFERENCES

1. M. K. Agoston, Algebraic Topology, *Pure and Applied Mathematics* **32**, Marcell Dekker, Inc., New York, Basel, 1976
2. D. W. Blackett, *Elementary Topology*, Academic Press, Inc., New York, 1982.
3. *Contests in Higher Mathematics*, 1949–1961, Akadémiai Kiadó, Budapest, 1968.
4. *Delta*; *Mathematical, Physical and Astronomical Popular Monthly* **115 / 7** (1983), Problem 235, p. 7.
5. D. Gale: The game of Hex and the Brouwer fixed-point theorem, *Amer. Math. Monthly* **86** (1979) 818–827.
6. J. E. Goodman, J. Pach, and C. K. Yap, Mountain climbing, ladder moving and the ring-width of a polygon, *Amer. Math. Monthly* **96** (1989) 494–510.
7. J. P. Huneke, Mountain Climbing, *Trans. Amer. Math. Soc.* **139** (1969) 383–391.
8. K. Kuratowski, *Introduction to set theory and topology*, International Series of Monographs on Pure and Applied Mathematics **13**, Pergamon Press, New York, 1961.
9. J. V. Whittaker, A mountain climbing problem, *Can. J. Math.* **18** (1966) 873–887.

VILMOS TOTIK was born in 1954 in Hungary. He is a professor at the Bolyai Institute, Szeged, Hungary and also at the University of South Florida, Tampa. He is a member of the Hungarian Academy of Sciences. He has (co)authored four research monographs: *Moduli of Smoothness*, *General Orthogonal Polynomials*, *Logarithmic Potentials with External Fields*, and *Weighted Approximation with Varying Weights*. His loves are his family (Veronika, Orsolya, Zoltán), his profession (approximation theory, orthogonal polynomials, potential theory), sports (soccer, tennis, table tennis), and any other normal activities (gardening, carpentry, ...).

Bolyai Institute, Szeged, Aradi v. tere, 1, 6720, HUNGARY and
University of South Florida, Tampa, FL 33620
totik@math.usf.edu

From the MONTHLY 50 years ago...

The following reports of Summer Sessions to be held in 1924 have been received.

University of Chicago, first term, June 16 to July 23; second term, July 24 to August 29. In addition to the usual courses in College algebra, Plane analytic geometry, and Calculus, the following advanced courses are announced: By Professor G. A. Bliss: Functions of a real variable; Thesis work in analysis. By Professor L. E. Dickson: Theory of Numbers, I; Thesis work in number theory. By Professor H. E. Slaught: Elliptic integrals; Differential equations. By Professor M. Fréchet: Theory of abstract sets; Theory of probability. By Professor E. T. Bell: General theory of numbers; Theory of equations. By Professor F. R. Moulton: Functions of infinitely many variables; Analytic mechanics, II. By Professor E. P. Lane: Synthetic projective geometry. By Doctor Mayme L. Logsdon, Introduction to higher algebra.

...Vol. 31 (1924) 211

A Mathematical Excursion: From the Three–door Problem to a Cantor-Type Set

Jaume Paradís, Pelegrí Viader, and Lluís Bibiloni

1. INTRODUCTION. We invite you, reader, on a mathematical trip. Our starting point is a well-known problem, the three-door problem (also known as the Monty Hall problem); our endeavors to solve it take us to a beautiful representation system for the real numbers in $(0,1]$ which, in turn, provides us with a nice enumeration of the positive rationals; as a bonus we can easily prove the irrationality of e. Our trip ends in the dark region of mysterious sets, where we find a simple description of a Cantor-type perfect set contained in $(0,1]$.

2. STARTING POINT: THE THREE-DOOR PROBLEM. Mathematics has always been enriched by a diversity of games and intellectual curiosities. These have provided an endless supply of problems that have acquired a life of their own, far removed from the recreational aspect of their origins. For example, the first building blocks of probability owe their existence to the analysis of gambling games carried out by Fermat and Pascal in the beginning of the XVIIth century. Undoubtedly Fermat himself was much attracted to mathematics thanks to Bachet's *Problèmes plaisants et délectables* of 1612 **[1]**, which was an introduction to Bachet's most famous book: the Latin translation of Diophantus' *Arithmetica*, in whose margins Fermat wrote the note that made his major theorem famous. Another important instance, E. Lucas' *Récréations Mathématiques* **[15]**, was a source of interesting problems at the beginning of the present century.

Let us start our excursion by setting a simple problem in the form of a seemingly innocent game.

2.1 The Three-Door Problem. In a TV contest, one of three shut doors hides a wonderful prize while the other two open onto a dismal void. The host proposes that the contestant choose one of the three doors. Then, as the contest rules establish, the host opens one of the other two doors wide showing the absence of any prize and offers the contestant the possibility of changing his/her choice. The contestant has to make a decision: to change or not to change.

Our mathematical challenge is to help the contestant make this decision by finding the probability of both possibilities. A (widely accepted) solution to the problem assigns probability $1/3$ to the option not to change and $2/3$ to the option to change. One way to reach this conclusion is the following reasoning:

The probability of choosing the right door in the first place is unquestionably $1/3$. *The probability that one of the other doors hides the prize is then* $2/3$. *If we choose not to change when we are offered the chance, our door still has the same probability of success,* $1/3$, *while now, the other door has a probability* $2/3$ *of hiding the prize.*

There are numerous references to this problem in the literature: see **[22]**, **[23]**, **[26]**, **[6]** or **[2]**.

2.2 A Reformulation of the Problem. Now tackle the same problem with a slightly different setting:

> In a TV contest, the host randomly hides a single prize in one of several boxes. The contestant chooses a box and then the host—who knows where the prize is—picks a box different from the one the contestant chose, opens it, and shows the empty contents to the contestant and the audience. The empty box is then discarded and, at this point, the contestant is permitted to choose a new, different, box, or may stick with the old one. In the first case, the contestant chooses a new box and holds it, while all remaining boxes, together with the one rejected by the contestant, are jumbled randomly. The same process continues until two boxes are left: the one the contestant holds and another one. The contestant is then offered the last possibility of change. After that, a mathematician, having followed the whole process attentively, says: "The contestant's probability of winning is $11/42$."
>
> A second mathematician, who has been fast asleep during the whole contest and does not know the initial number of boxes, but is familiar with the rules of the show, wakes up, hears the last utterance, and says: "From what my colleague says I deduce that initially there were 7 boxes and the contestant changed on two occasions, when there were 4 and 3 boxes to choose from."
>
> How did the two mathematicians reach their conclusions?

We call the preceding reformulation and generalization of the three-door problem the *n-box* problem; boxes are more suitable than doors when it comes to jumbling them randomly.

2.3 A Hint. We suggest our reader try to find the probability of each of the 2^{n-2} possible strategies that our contestant can follow, if n is the initial number of boxes in the game. There are 2^{n-2} because any set of choices can be described as a string of 1 and 0 (1 for changing and 0 for sticking) and there are $n-2$ offers of change.

A strategy can be represented by a strictly decreasing sequence of positive integers $\{n, a_k, a_{k-1}, \dots, a_2, a_1\}$ such that $n > a_k > a_{k-1} > \cdots > a_2 > a_1 \geq 1$, where a_i denotes that a change of boxes was made when there were a_i boxes to choose from (notice that $a_k \neq n-1$).

2.4 The Solution to the Problem. If no change whatsoever is made, the probability of winning is obviously $1/n$. If a last-minute change is made (when there is only one box offered besides that initially chosen), the probability of winning is $(n-1)/n$.

If we describe any other strategy by the convention described in the preceding hint, the first change is made when the contestant can choose from a_k boxes. The probability of choosing the right box is the probability of having previously chosen the wrong one times the probability of choosing correctly among a_k boxes, that is:

$$p_k = \left(1 - \frac{1}{n}\right)\frac{1}{a_k}.$$

For the contestant's next change, the same reasoning shows that

$$p_{k-1} = (1 - p_k)\frac{1}{a_{k-1}} = \left(1 - \left(1 - \frac{1}{n}\right)\frac{1}{a_k}\right)\frac{1}{a_{k-1}},$$

which can be expressed as

$$p_{k-1} = \frac{1}{a_{k-1}} - \frac{1}{a_{k-1} \cdot a_k} + \frac{1}{a_{k-1} \cdot a_k \cdot n}.$$

Iterating the process, we have for the last change

$$p_1 = (1 - p_2)\frac{1}{a_1} = \frac{1}{a_1} - \frac{1}{a_1 \cdot a_2} + \cdots + \frac{(-1)^{k-1}}{a_1 \cdot a_2 \cdots a_k} + \frac{(-1)^k}{a_1 \cdot a_2 \cdots a_k \cdot n}. \quad (2.1)$$

In (2.1) we have $1 \le a_1 < a_2 < \cdots < a_k < n - 1$. A strategy is described by a subset of $\{1, 2, \ldots, n-2\}$; ($\varnothing$ corresponds to the strategy of making no change at all). For each strategy $\{a_1, a_2, \ldots, a_k\} \subset \{1, 2, \ldots, n-2\}$, the probability of winning is given by (2.1).

This accounts for our first mathematician's assertion, as

$$n = 7,\ k = 2,\ a_2 = 4,\ a_1 = 3, \quad \text{and} \quad p_1 = \frac{1}{3} - \frac{1}{3 \cdot 4} + \frac{1}{3 \cdot 4 \cdot 7} = \frac{11}{42}.$$

To help understand our second mathematician's claim, we can play a little with what we have and find a few more probabilities in the case $n = 7$. A patient completion of Table 1 (2^5 entries) would show a most interesting fact: different strategies correspond to different probabilities. If similar tables for other values of n were made, we would notice that in all cases the probabilities obtained were different, not only within one table but also among different tables. This motivates the following result.

Theorem 1. *Any rational number $p/q \in (0, 1]$ has a unique representation*

$$\frac{p}{q} = \frac{1}{a_1} - \frac{1}{a_1 \cdot a_2} + \cdots + \frac{(-1)^{k-1}}{a_1 \cdot a_2 \cdots a_k}, \quad (2.2)$$

where a_i are positive integers such that

$$1 \le a_1 < a_2 < \cdots < a_{k-1} < a_k - 1.$$

TABLE 1

Strategy	Probability
$\{1,2,3,4,5\}$	$1 - \frac{1}{2} + \frac{1}{2 \cdot 3} - \frac{1}{2 \cdot 3 \cdot 4} + \frac{1}{2 \cdot 3 \cdot 4 \cdot 5} - \frac{1}{2 \cdot 3 \cdot 4 \cdot 5 \cdot 7} = \frac{62}{105}$
$\{1,2,3,4\}$	$1 - \frac{1}{2} + \frac{1}{2 \cdot 3} - \frac{1}{2 \cdot 3 \cdot 4} + \frac{1}{2 \cdot 3 \cdot 4 \cdot 7} = \frac{53}{84}$
$\{1,2,3\}$	$1 - \frac{1}{2} + \frac{1}{2 \cdot 3} - \frac{1}{2 \cdot 3 \cdot 7} = \frac{9}{14}$
$\{2,4,5\}$	$\frac{1}{2} - \frac{1}{2 \cdot 4} + \frac{1}{2 \cdot 4 \cdot 5} - \frac{1}{2 \cdot 4 \cdot 5 \cdot 7} = \frac{111}{280}$

Proof: The interval $(0,1]$ can be expressed as the disjoint union $(0,1] = \bigcup_{n=1}^{\infty}((n+1)^{-1}, n^{-1}]$, so any $\alpha \in (0,1]$ belongs to one of the intervals $((n+1)^{-1}, n^{-1}]$. Consequently,

$$\alpha = \frac{1}{n} - \lambda_1\left(\frac{1}{n} - \frac{1}{n+1}\right) = \frac{1}{n} - \frac{\lambda_1}{n(n+1)},$$

for some $\lambda_1 \in [0,1)$. If we denote $\alpha_1 = \lambda_1/(n+1)$, we have $\alpha = (1-\alpha_1)/n$ and $\alpha_1 \in (0,(n+1)^{-1})$. Applying the same procedure to α_1 we get $\alpha_1 = (1-\alpha_2)/m$ $(m > n)$, and we eventually get an expansion of the form (2.2):

$$\alpha = \frac{1}{n} - \frac{1}{n \cdot m} \cdot \alpha_2 \qquad (m > n).$$

The algorithm that leads to (2.2) can be summarized by iterating the two operations

$$\begin{aligned} a_i &= \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor \qquad \text{with} \qquad \alpha_0 = \alpha, \\ \alpha_i &= 1 - \alpha_{i-1} \cdot a_i, \end{aligned} \tag{2.3}$$

where $\lfloor x \rfloor$ denotes the greatest integer less or equal than x.

If α is irrational, all the α_i are irrational and the algorithm never terminates. This proves more than promised in the phrasing of the theorem; it proves the existence of an infinite expansion of the form (2.2) for any irrational in $(0,1]$.

If $\alpha = p/q$ is a rational number in lowest terms, the algorithm becomes a modified Euclidean algorithm. If we divide q by p:

$$q = a_1 \cdot p + r_1 \qquad (0 \le r_1 < p),$$

it is obvious that

$$\left\lfloor \frac{q}{p} \right\rfloor = a_1 \qquad \text{and} \qquad \frac{r_1}{q} = \alpha_1.$$

Next we would perform the division of q by r_1:

$$q = a_2 \cdot r_1 + r_2 \qquad (0 \le r_2 < r_1),$$

and so on. Since the sequence of remainders r_i is strictly decreasing $(p > r_1 > r_2 > \cdots)$, the algorithm eventually terminates with $r_k = 0$. Therefore the expansion (2.2) is finite and the last two divisions are

$$\begin{aligned} q &= a_{k-1} \cdot r_{k-2} + r_{k-1} \qquad (0 \le r_{k-1} < r_{k-2}) \\ q &= a_k \cdot r_{k-1}. \end{aligned}$$

Thus, $a_{k-1} \cdot r_{k-2} = (a_k - 1) \cdot r_{k-1}$, and since $r_{k-1} < r_{k-2}$ we have $a_{k-1} < a_k - 1$.

The uniqueness of the expansions comes from the double inequality

$$\frac{1}{a_1+1} < \frac{1}{a_1} - \frac{1}{a_1 \cdot a_2} + \cdots \le \frac{1}{a_1}.$$

The only duplicate expansion is obtained in the finite case, due to the equality

$$\frac{1}{n+1} = \frac{1}{n} - \frac{1}{n(n+1)}.$$

That is the reason for the exclusion of two consecutive integers at the end of the expansion. ■

We denote the expansion (2.2) by

$$\langle a_1, a_2, \dots \rangle.$$

Now, we see how our second mathematician could reconstruct the events of the contest. Starting with the probability 11/42, we compute $42 = 11 \cdot \mathbf{3} + 9 = 9 \cdot \mathbf{4} + 6 = 6 \cdot \mathbf{7}$. Thus

$$\frac{11}{42} = \frac{1}{3} - \frac{1}{3 \cdot 4} + \frac{1}{3 \cdot 4 \cdot 7} = \langle 3, 4, 7 \rangle.$$

The uniqueness of the expansion allows the second mathematician to say:

There were 7 *boxes and the contestant changed on two occasions, when there were* 4 *and* 3 *boxes to choose from.*

3. A REPRESENTATION SYSTEM FOR THE REAL NUMBERS IN (0, 1]. We have solved our generalized n-door problem and, at the same time, have discovered a system of representation for the real numbers α in (0, 1]:

$$\alpha = \frac{1}{a_1} - \frac{1}{a_1 a_2} + \frac{1}{a_1 a_2 a_3} - \cdots$$

where $1 \le a_1 < a_2 < a_3 < \cdots$.

The first mathematicians who paid any attention to these expansions were Lambert (1770) and Lagrange (1798); see **[13]** and **[12]**. Later, Ostrogradsky (†1860), and Sierpiński (1911) were the first to develop a few of their numerical properties; see **[19]** and **[27]**. Pierce (1929) used the model in an algorithm to find algebraic roots of polynomials, **[18]**. Some authors have attached Pierce's name to the expansion that had been previously referred to as "Lambert fractions" or "ascending fractions." In a 1986 presentation, Shallit **[24]** studied the metric theory of the model following the methods used for the non-alternated expansions (Engel's series) developed in 1947 by Borel **[3]** and Lévy **[14]**, and later by Erdős, Rényi, and Szüsz **[4]**, improved by Rényi in 1962 **[20]**. There is also a 1987 paper by A. Knopfmacher and J. Knopfmacher **[10]**, who use the model to construct the real numbers. Some interesting new results related to Pierce expansions can be found in **[25]** and **[11]**.

The infinite Pierce expansion $\langle 1, 2, 3, 4, \dots \rangle$ is the Taylor expansion of $1 - e^x$ for $x = -1$:

$$1 - \frac{1}{e} = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots = \langle 1, 2, 3, \dots \rangle.$$

Incidentally, this proves the irrationality of e. Other examples are not so obvious:

$$\langle 1, 3, 5, 7, \dots \rangle = \frac{1}{\sqrt{e}} \sum_{n=0}^{\infty} \frac{1}{2^n \cdot n! \cdot (2n+1)}.$$

As a system of representation, Pierce expansions are not bad: Truncating the expansion of $\alpha = \langle a_1, a_2, \dots, a_n \rangle$ at level n, provides quite a good approximation to α:

$$|\alpha - \langle a_1, a_2, \dots, a_n \rangle| < \frac{1}{a_1 \cdots a_n \cdot a_{n+1}}, \tag{3.1}$$

which, in the worst case ($a_i = i$, $i = 1, 2, \dots$), is of the order $1/(n+1)!$.

4. A NEW ENUMERATION FOR THE POSITIVE RATIONALS. The most famous enumeration of the positive rationals is the diagonal ordering

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{3}{1}, \frac{2}{2}, \frac{1}{3}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{5}{1}, \frac{4}{2}, \frac{3}{3}, \frac{2}{4}, \frac{1}{5}, \dots$$

All fractions appear in this scheme, repeated infinitely many times; p/q appears in position $(1/2)(p+q-1)(p+q-2)+q$. After suppressing repetitions, to determine the position of the irreducible ones is, as Prof. Hardy says in [7, p. 1], *more complicated*: the computational complexity of the diagonal ordering algorithm, if one suppresses all repetitions, is exponential. This problem is intimately related to the representation of rational numbers **[17]**.

The basic idea is to use the binary representation of a positive integer n as a string of 0 and 1. Some of these strings can be considered strategies in our generalized n-box problem. We use the corresponding Pierce expansion to assign a rational to our n.

To any strictly increasing finite sequence of positive integers $\{a_1, a_2, \dots, a_k\}$ with $1 \le a_1 < a_2 < \cdots < a_k$, we associate the positive integer $n = 2^{a_1-1} + 2^{a_2-1} + \cdots + 2^{a_k-1}$, or, what amounts to the same, the number n that in the binary system is written, from right to left, as 1 in positions $a_1, a_2, \dots, a_k$ and 0 elsewhere. For example,

$$\{1,3,5,8\} \longrightarrow 2^0 + 2^2 + 2^4 + 2^7 = 10010101.$$

Now, to any rational number $p/q \in (0,1]$ we associate its Pierce expansion $\langle a_1, a_2, \dots, a_k \rangle$, which may be regarded as the strictly increasing finite sequence of positive integers $\{a_1, a_2, \dots, a_k\}$, where $a_k > 1 + a_{k-1}$. Its corresponding positive integer has the binary form $10\dots$, with a 0 in the next-to-last position as we go from right to left. To any rational number $q/p > 1$, we associate the Pierce expansion corresponding to its inverse $p/q = \langle a_1, a_2, \dots, a_k \rangle$ and we then consider the strictly increasing finite sequence of positive integers $\{a_1, a_2, \dots, a_{k-1}, a_k - 1, a_k\}$. Its corresponding positive integer has the binary form $11\dots$, with a 1 in the next-to-last position as we go from right to left.

Conversely, to any positive integer n written in the binary system as

$$2^{a_1} + 2^{a_2} + \cdots + 2^{a_k} \qquad \text{with} \qquad 0 \le a_1 < a_2 < \dots < a_k,$$

we assign the rational number

$$\langle 1+a_1, 1+a_2, \dots, 1+a_k \rangle \in (0,1] \quad \text{if } a_k > 1 + a_{k-1},$$

$$\frac{1}{\langle 1+a_1, 1+a_2, \dots, 1+a_{k-2}, 1+a_k \rangle} > 1 \quad \text{if } a_k = 1 + a_{k-1}.$$

The uniqueness of the Pierce expansion of any rational number in $(0,1]$ ensures the bijectivity of the map just defined between the positive integers and the positive rationals.

An example may help us understand the map we have just defined. Let us find what rational occupies place 10^{12} in our enumeration. First we write 10^{12} in binary:

$$10^{12} = 2^{39} + 2^{38} + 2^{37} + 2^{35} + 2^{31} + 2^{30} + 2^{28} + 2^{26} + 2^{23} + 2^{21} + 2^{18} + 2^{16} + 2^{12},$$

which corresponds to the fraction (in this case a rational greater than 1):

$$\frac{1}{\langle 13,17,19,22,24,27,29,31,32,36,38,40\rangle} = \frac{94232197736202240}{6843703050416119}.$$

The algorithm to find the fraction occupying a given place n has a polynomial computational complexity. The inverse algorithm would also have polynomial complexity, assuming the correctness of a conjecture formulated by Erdős and Shallit in [5] concerning the upper bound of the length of the finite Pierce expansion of p/q:

$$\text{length of the Pierce expansion of } \frac{p}{q} = O\big((\log q)^2\big).$$

5. A CLOSER LOOK AT PIERCE EXPANSIONS. Let us contemplate what we have accomplished and examine Pierce expansions more closely. For each real number α in (0, 1], define its i-th projection $\omega_i(\alpha)$ to be the map that assigns to α its i-th partial quotient: if $\alpha = \langle a_1, a_2, a_3, \dots \rangle$, then $\omega_i(\alpha) = a_i$.

A *cylinder* of order k is the set of numbers such that the first k partial quotients are fixed:

$$C(a_1, a_2, \dots, a_k) = \{\alpha \in (0,1] : \omega_1(\alpha) = a_1, \omega_2(\alpha) = a_2, \dots, \omega_k(\alpha) = a_k\}.$$

A cylinder of any order is an interval of length

$$|C(a_1, a_2, \dots, a_k)| = \frac{1}{a_1 \cdot a_2 \cdots a_k \cdot (1 + a_k)}.$$

Moreover, a cylinder of order k is the disjoint union of all the cylinders of order $k + 1$ contained in it:

$$C(a_1, a_2, \dots, a_k) = \bigcup_{j=1+a_k}^{\infty} C(a_1, a_2, \dots, a_k, j).$$

We can also consider generalized cylinders, in which the fixed partial quotients are not the first k; they are not intervals, though they are still unions of intervals. The simplest is

$$H[\omega_k(\alpha) = n] = \{\alpha \in (0,1] : \omega_k(\alpha) = n\},$$

which is a union of cylinders:

$$H[\omega_k(\alpha) = n] = \bigcup_{1 \le a_1 < a_2 < \cdots < a_{k-1} \le n-1} C(a_1, a_2, \dots, a_{k-1}, n).$$

Consequently, its Lebesgue measure is:

$$\begin{aligned} \lambda(H[\omega_k(\alpha) = n]) &= \sum_{1 \le a_1 < a_2 < \cdots < a_{k-1} \le n-1} \frac{1}{a_1 a_2 \cdots a_{k-1} n(n+1)} \\ &= \frac{1}{n(n+1)} \sum_{1 \le a_1 < a_2 < \cdots < a_{k-1} \le n-1} \frac{1}{a_1 a_2 \cdots a_{k-1}}. \end{aligned} \tag{5.1}$$

One way to evaluate the last sum in (5.1) is to multiply inside by $(n-1)!$ and divide outside by the same quantity:

$$\sum_{1 \le a_1 < \cdots < a_{k-1} \le n-1} \frac{1}{a_1 a_2 \cdots a_{k-1}} = \frac{1}{(n-1)!} \sum_{1 \le a_1 < \cdots < a_{n-k} \le n-1} a_1 a_2 \cdots a_{n-k}. \tag{5.2}$$

The right-hand sum in (5.2) can be viewed as the coefficient of x^k in the polynomial $x(x+1)(x+2)\cdots(x+n-1)$, which is a *Stirling number of the second kind*. The properties of Stirling numbers (both of the first and the second kind) can be found in [**8**, pp. 243–253], whose notation we follow: $\begin{bmatrix} n \\ k \end{bmatrix}$. Thus,

$$x(x+1)(x+2)\cdots(x+n-1) = \begin{bmatrix} n \\ 1 \end{bmatrix} x + \begin{bmatrix} n \\ 2 \end{bmatrix} x^2 + \cdots + \begin{bmatrix} n \\ n \end{bmatrix} x^n. \tag{5.3}$$

Using this notation, the sums in (5.2) are equal to $\begin{bmatrix} n \\ k \end{bmatrix}/(n-1)!$. Finally, we have

$$\lambda(H[\omega_k(\alpha) = n]) = \frac{1}{n(n+1)} \cdot \frac{\begin{bmatrix} n \\ k \end{bmatrix}}{(n-1)!} = \frac{\begin{bmatrix} n \\ k \end{bmatrix}}{(n+1)!}. \tag{5.4}$$

With (5.4) and a very simple property of Stirling numbers:

$$\begin{bmatrix} n \\ 1 \end{bmatrix} + \begin{bmatrix} n \\ 2 \end{bmatrix} + \cdots + \begin{bmatrix} n \\ n \end{bmatrix} = n!$$

which follows from (5.3) by considering $x = 1$, it is easy to prove the following:

Theorem 2. *The set of real numbers in* $(0,1]$ *whose Pierce expansion contains the integer n has Lebesgue measure* $1/(n+1)$.

Theorem 2 has an immediate corollary:

Theorem 3. *The set of real numbers in* $(0,1]$ *whose Pierce expansion does not contain the integer n has Lebesgue measure* $n/(n+1)$.

It is not difficult to generalize Theorem 3:

Ttheorem 4. *The set of real numbers in* $(0,1]$ *whose Pierce expansion does not contain the distinct integers m and n has Lebesgue measure* $nm/(n+1)(m+1)$.

All these results can be found in [**24**].

6. A CANTOR-TYPE PERFECT SET. A set in **R** that is closed and has no isolated points is said to be *perfect*. Such a set coincides with the set of its limit-points (its *derived* set). The easiest example of a perfect set in **R** is a closed interval, but there are perfect sets that not only are not intervals, they do not even contain any interval. A classic example of this behavior is Cantor's ternary set

$$\mathscr{C} = \left\{ \alpha \in [0,1] : \alpha = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, \quad \text{with } a_i \in \{0,2\} \right\},$$

that is to say, the set of all real numbers in [0, 1] that can be written in the ternary system without the digit 1. Geometrically, Cantor's set can be described by

iterating indefinitely the following procedure: From $[0,1]$ we suppress the central open interval $(1/3, 2/3)$ and from the remaining two intervals, we suppress the corresponding central open intervals $(1/9, 2/9)$ and $(7/9, 8/9)$, and so on. The points that remain after the suppression of all these open intervals constitute Cantor's set.

Cantor's set is a perfect set; it is uncountable (a consequence of being perfect) and has measure zero (its complement in $[0,1]$ is the union of countably many disjoint open intervals with total length 1). The interior of $\mathscr{C}$ is obviously empty. See **[21]** for details.

Now, let us consider the set C of real numbers in $(0,1]$ whose Pierce expansion contains no odd integers. According to Theorem 4, the Lebesgue measure of C is

$$\lambda(C) = \prod_{n=1}^{\infty}\left(1 - \frac{1}{2n}\right) = 0.$$

The set C is uncountable since we can establish a one-to-one correspondence between its elements and $(0,1]$:

$$\langle a_1, a_2, \ldots, a_n, \ldots \rangle \leftrightarrow \langle \frac{a_1}{2}, \frac{a_2}{2}, \ldots, \frac{a_n}{2}, \ldots \rangle.$$

It is also easy to prove that C, like Cantor's set, is perfect and its interior is empty.

7. A CANTOR-TYPE PERFECT SET OF TRANSCENDENTAL NUMBERS. In 1851 J. Liouville established a very important result that permitted him to exhibit, for the first time in mathematics, a transcendental real number (a real number that is not the root of any polynomial equation with rational coefficients). A real number α is said to be algebraic of degree n if there is a polynomial of degree n (but not lower) with rational coefficients that has α as a root.

Liouville's Theorem. [**16**, pp. 87–93] *If α is algebraic of degree n, $(n > 1)$, there exists a constant M (depending on α) such that*

$$\left|\alpha - \frac{a}{b}\right| \geq \frac{M}{b^n}.$$

for all rational numbers a/b.

Consider the following Pierce expansion

$$l_p = \langle p^{2!}, p^{3!-2!}, \ldots, p^{n!-(n-1)!}, \ldots \rangle = \frac{1}{p^{2!}} - \frac{1}{p^{3!}} + \frac{1}{p^{4!}} - \cdots + \frac{(-1)^n}{p^{n!}} + \cdots,$$

where p is any positive integer.

The number l_p is transcendental because (3.1) tells us that

$$\left| l_p - \left(\frac{1}{p^{2!}} - \frac{1}{p^{3!}} + \cdots + \frac{(-1)^k}{p^{k!}} \right) \right| = \left| l_p - \frac{a}{p^{k!}} \right| < \frac{1}{p^{(k+1)!}} < \frac{1}{\left(p^{k!}\right)^k}, \quad (7.1)$$

which would contradict Liouville's theorem if l_p were algebraic of degree k.

We may now consider the set L_p of all real numbers in $(0,1]$ whose Pierce expansion contains only integers extracted from the Pierce expansion of l_p. It is at

once seen that L_p has measure zero, is uncountable, and (since all its elements satisfy inequalities like (7.1)) consists entirely of transcendental numbers.

8. LOOKING BACK. We have started with an interesting and controversial problem, the Monty Hall dilemma (which has a totally probabilistic set-up), and have reached some very peculiar subsets of $[0, 1]$: uncountable closed sets with an empty interior, without isolated points and of measure zero—the same structure as Cantor's ternary set, with the added feature of being formed exclusively by transcendental numbers. The connection between such apparently distant concepts is a beautiful system for real number representation, Pierce expansions, which *exactly* describe the probability of each one of the possible strategies that can be followed by the contestant in a generalization of the Monty Hall problem: the n-box problem. We have also encountered a nice (and new) enumeration of the positive rationals that is based on both the strategies in the n-box problem and Pierce expansions.

Undoubtedly we have overlooked many unexplored places, but we hope you have enjoyed the few we have been lucky enough to find.

REFERENCES

1. C. Bachet de Méziriac, *Problèmes plaisants et délectables*, 1st ed., 1612. 3rd ed., Gauthier–Villars, Paris, 1874.
2. E. Barbeau, Fallacies, flaws, and flimflam: The problem of the car and the goats, *College Math. J.* **24** (1993) 149–154.
3. E. Borel, Sur les développements unitaires normaux, *Comptes Rendus* **225** (1947) 74–79.
4. P. Erdős, A. Rényi, and P. Süsz, On Engel's and Sylvester's series, *Annales Univ. Sci. Budapest, Sectio Math.* **1** (1958) 7–32.
5. P. Erdős and J. O. Shallit, New bounds on the Length of Finite Pierce and Engel Series, *Séminaire de Théorie des nombres de Bordeaux* **3** (1991) 43–53.
6. L. Gillman, The car and the goats, *Amer. Math. Monthly* 99 (1992) 3–7.
7. G. H. Hardy, *A Course of Pure Mathematics*, Cambridge Univ. Press, 8th ed., London, 1941.
8. R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, 6th ed., Reading, Mass., 1990.
9. A. Ya. Khintchine, *Continued Fractions*. Translated by Peter Wynn. P. Noordhoff, Ltd., Groningen, 1963.
10. A. Knopfmacher and J. Knopfmacher, Two constructions of the real numbers via alternating series, *Internat. J. Math. and Math. Sci.* **12** (1989) 603–613.
11. A. Knopfmacher and M. E. Mays, Pierce expansions of ratios of Fibonacci and Lucas numbers and polynomials, *Fibonacci Quarterly* **33** (1995) 153–163.
12. J. L. Lagrange, Essai d'analyse numérique sur la transformation des fractions, *Journal de l'École Polytechnique*, Vème cahier, t. II, prairial, an VI. Also in *Oeuvres* ed. J. A. Serret, Tome VII, Section quatrième, 291–313.
13. J. H. Lambert, *Verwandlung der Brüche*: *Beiträge zum Gebrauche der Mathematik und deren Anwendung*, Part 2. Berlin 1770.
14. P. Lévy, Remarques sur un théoreme de M. Émile Borel, *Comptes Rendus* **225** (1947) 918–919.
15. E. Lucas, *Récréations mathématiques*, Gauthier–Villars, Paris, 1891–1894. Reprinted by A. Blanchard, Paris, 1960.
16. I. Niven, *Irrational Numbers*, Carus Math. Monographs, **11**. Wiley, NY, 1956.
17. J. Paradís, Ll. Bibiloni, and P. Viader, On Actually Computable Bijections between **N** and $\mathbb{Q}^+$, *Order* **13** (1996) 369–377.
18. T. A. Pierce, On an Algorithm and its Use in approximating Roots of Algebraic Equations, *Amer. Math. Monthly* **36** (1929) 532–525.
19. E. Ya. Remez, On series with alternating signs, which may be related to two algorithms of M. V. Ostrogradski for the approximation of irrationals numbers, *Matem. Nauk.*, (N.S.) 6,5 **45** (1951) 33–42.
20. A. Rényi, A new Approach to the theory of Engel's series, *Annales Univ. Sci. Budapest, Sectio Math.* **5** (1962) 25–32.

21. W. Rudin, *Principles of Mathematical Analysis*, McGraw–Hill, New York, 3d ed. 1976.
22. S. Selvin, A problem in probability, *Amer. Statist.* **29** (1975) 67.
23. S. Selvin, On the Monty Hall problem, *Amer. Statist.* **29** (1975) 134.
24. J. O. Shallit, Metric Theory of Pierce Expansions, *Fibonacci Quarterly* **24** (1986) 22–40.
25. J. O. Shallit, Pierce Expansions and Rules for the Determination of Leap Years, *Fibonacci Quarterly* **32** (1994) 416–423.
26. M. Shaughnessy, T. Dick, Monty's dilemma: should you stick or switch?, *Math. Teacher* **84** (1993) 252–256.
27. W. Sierpiński, Sur quelques algorithmes pour développer les nombres réels en séries, (1911). In *Oeuvres choisies*, Tome I, Warszawa, 1974, 236–254.

JAUME PARADÍS graduated in Mathematics at the University of Barcelona and received his Ph.D. in Mathematics at the Universitat Politècnica de Catalunya, under the supervision of Professor J. Egozcue. He now has tenure at the Universitat Pompeu Fabra, Barcelona. His early research includes several books and papers on the History of Algebra, mainly on Viète's and Fermat's contributions. His present research is focused on Pierce expansions and other systems of real number representation and on singular functions defined with their help.
Applied Mathematics, Univ. Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain
jaume.paradis@econ.upf.es

PELEGRÍ VIADER is Associate Professor of Mathematics at the Universitat Pompeu Fabra, Barcelona. He received his Ph.D. in mathematics at the Universitat Politècnica de Catalunya. His main research interests are Pierce expansions and other algorithms of real number representation, singular functions and History of Mathematics.
Applied Mathematics, Univ. Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain
pelegri.viader@econ.upf.es

LLUÍS BIBILONI studied Mathematics at Universitat de Barcelona. He became involved in teacher training, in the Departament de Didàctica de les Matemàtiques i de les Ciències Experimentals, at the Universitat Autònoma de Barcelona where he is professor of Mathematics Education. He received his Ph.D. in mathematics at the Universitat Politècnica de Catalunya. His interests include the Philosophy and History of Science and of Mathematics, and has been deeply influenced by the work of George Pólya on problem solving whose ideas is trying to pursue at a practical teaching level and at research level in the field of Mathematics Education. He is also interested in computer–aided teaching. His favorite mathematical subjects are Number Theory, Analysis, Mathematical Logic, Nonstandard Analysis, Mathematical Physics, and Probability Theory.
Facultat de Ciències de l'Educació, Univ. Autònoma Barcelona, 08193, Bellaterra, Barcelona, Spain.
l.bibiloni@uab.es

NOTES

Edited by **Jimmie D. Lawson and William Adkins**

An Inequality Relating the Circumradius and Diameter of Two-Dimensional Lattice-Point-Free Convex Bodies

Poh Wah Awyong

1. Introduction. The first ideas of convex sets date as far back as Archimedes but it was not until the end of the last century that a systematic study gave rise to the subject as an independent branch in mathematics. In particular, many geometric inequalities for convex bodies have been obtained; see **[1]**, **[8]**, **[10]**, **[11]**, and **[12]**.

At the turn of the century, Minkowski **[6]** published his famous Convex Body Theorem, which is the basis for the geometry of numbers. The idea is to interpret integer solutions of equations or inequalities as points with integer coordinates (lattice points). Minkowski's work provides the link between the theory of convex sets and the geometry of numbers. Minkowski's Theorem states that if a convex set in the plane is symmetric about the origin and its interior contains no other lattice point, then its area is at most 4. By studying other geometrical functionals defined on a convex set and varying the conditions on Minkowski's Theorem, many inequalities may be obtained for lattice constrained sets; see **[2]**, **[3]**, **[4]**, **[5]**, and **[9]**.

In this note, we prove an inequality concerning the circumradius and diameter of a planar convex set. We use this inequality to obtain a corresponding result for a lattice-point-free convex set.

2. Notation and Definitions. Throughout this note, K denotes a compact, convex set in the plane. The *circumradius* of K, denoted by $R(K) = R$, is the radius of the smallest disk containing K. The *inradius* of K, denoted by $r(K) = r$, is the radius of the largest disk contained in K. The diameter of K, denoted by $D(K) = D$, is the maximal distance between any two points of K. The *width of K taken in a particular direction* is the distance between the two parallel tangents to K perpendicular to the given direction. The *width* of K denoted by $w(K) = w$, is the minimum of widths taken in all directions.

3. Motivation. The result in this note is motivated by an inequality by Blaschke, which states $w \leq 3r$ for any planar convex set K, with equality when and only when K is an equilateral triangle **[11**, p. 18**]**. This inequality may be rewritten as

$$w - 2r \leq \frac{w}{3}. \tag{1}$$

If K contains no interior lattice point, we have the following result by Scott **[7]**:

$$w \leq \frac{1}{2}(2 + \sqrt{3}), \tag{2}$$

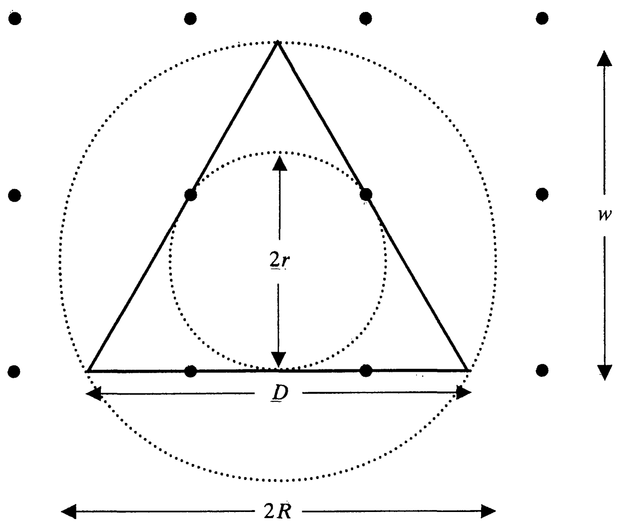


Figure 1. Equilateral triangle with no interior lattice points

with equality when and only when K is congruent to the equilateral triangle shown in Figure 1.

Combining (2) with (1), we have

$$w - 2r \leq \frac{1}{6}(2 + \sqrt{3}), \tag{3}$$

with equality when and only when K is congruent to the equilateral triangle shown in Figure 1.

In this note, we prove 'duals' of (1) and (3):

Theorem. *Let K be a planar, compact, convex set. Then*

$$2R - D \leq \frac{2}{3}(2 - \sqrt{3})w, \tag{4}$$

with equality when and only when K is an equilateral triangle. If K contains no lattice point in its interior, then

$$2R - D \leq \frac{1}{3}, \tag{5}$$

with equality when and only when K is congruent to the equilateral triangle shown in Figure 1.

4. Proof of the Theorem. We may assume that the interior of K is nonempty, otherwise, either $K = \varnothing$ or K is a line segment. If $K = \varnothing$, then (4) is trivially true. If K is a line segment then $D = 2R$, $w = 0$, and again, (4) is trivially true. Hence we may assume that $r \neq 0$. It follows that $w \neq 0$. We now define and seek to maximize the functional

$$f(K) = \frac{1}{w(K)}(2R(K) - D(K)) = \frac{1}{w}(2R - D).$$

Clearly, $f(K) \geq 0$ since $D \leq 2R$. We first recall that the circumcircle of a set K either contains two diametrically opposite points of K or else it contains three

points on the boundary of K that form the vertices of an acute-angled triangle [**11**, p. 59]. In the first case, $2R = D$ and $f(K) = 0$, so K is not maximal. Hence we may assume that K contains an acute-angled triangle T with $R(T) = R(K)$. Furthermore, since T is contained in K, $D(T) \leq D(K)$ and $w(T) \leq w(K)$. It follows that $f(K) \leq f(T)$. Hence it suffices to maximize $f(K)$ for acute-angled triangles T.

Let $T = \Delta XYZ$ be an acute-angled triangle with $\angle Y \leq \angle X \leq \angle Z$, as shown in Figure 2. Since $\angle Z$ is the largest angle, it follows that $XY = D$. We first apply to T

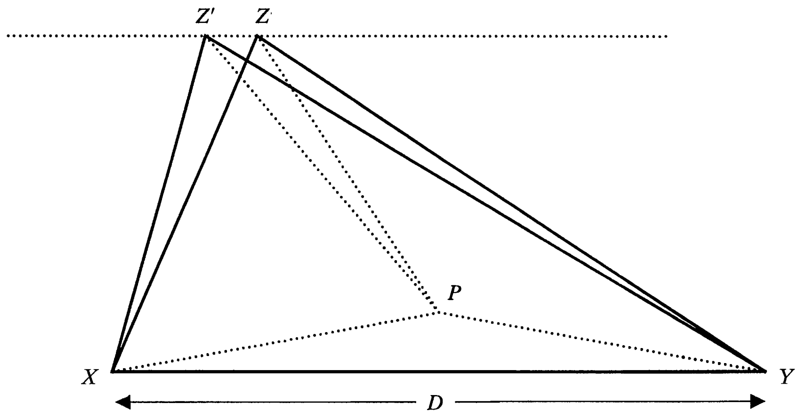


Figure 2. Shear applied to the triangle T

a shear parallel to XY to obtain the triangle $T' = XYZ'$ with $YZ' = XY = D$. Let P and P' be the circumcentres of T and T' respectively. Since P and P' both lie on the perpendicular bisector of the line segment XY, and since $PZ' > PZ = R(T)$, it follows that P' is farther away from XY than the point P. Hence $R(T') > R(T)$. Furthermore $D(T') = D(T)$ and $w(T') = w(T)$. It follows that $f(T') \geq f(T)$. Hence we need consider only those cases for which T is an isosceles triangle with vertex angle at Y. In this case $\angle X = \angle Z = \alpha \geq \pi/3$.

We note that $w = D \sin 2\alpha$ and the sine rule gives $2R = D/\sin \alpha$. Hence we have

$$f(k) = \frac{1}{w}\left(\frac{1}{\sin \alpha} - 1\right)D = \left(\frac{1}{\sin \alpha} - 1\right)\left(\frac{1}{\sin 2\alpha}\right).$$

Letting $t = \tan \alpha$ gives

$$\begin{aligned} f(K) &= \left(\frac{\sqrt{1+t^2}}{t} - 1\right)\left(\frac{1+t^2}{2t}\right) = \frac{1}{2}\left(\sqrt{1+t^2} - t\right)\left(\frac{1+t^2}{t^2}\right) \\ &= \frac{1}{2}\left(\sqrt{1+t^2} - t\right)\left(\frac{1}{t^2} + 1\right) = \frac{1}{2}g(t)h(t). \end{aligned}$$

We note that

$$g(t) = \sqrt{1+t^2} - t > 0, \qquad g'(t) = \frac{1}{\sqrt{1+t^2}} - 1 < 0,$$

$$h(t) = \frac{1}{t^2} + 1 > 0, \qquad h'(t) = -\frac{2}{t^3} < 0.$$

Since $f(K)$ is a product of positive, decreasing functions of t, it is itself a positive, decreasing function of t. Since $\alpha \geq \pi/3$, we have $t \geq \sqrt{3}$. Hence the maximal value of $f(K)$ is attained when $t = \sqrt{3}$, that is, when T is an equilateral triangle. In this case

$$f(K) = \frac{1}{w}(2R - D) \leq \frac{2}{3}(2 - \sqrt{3}).$$

Now suppose that K has no lattice point in its interior. Combining (4) with (2) gives

$$2R - D \leq \frac{1}{3},$$

with equality when and only when K is congruent to the equilateral triangle shown in Figure 1. ∎

REFERENCES

1. T. Bonnesen and W. Fenchel, *Theory of Convex Bodies*. Translated and edited by L. Boron, C. Christenson, and B. Smith, with the collaboration of W. Fenchel, BCS Associates, Moscow, Idaho, 1987.
2. P. Erdős, P. M. Gruber, and J. Hammer, *Lattice points*, Longman, Essex, 1989.
3. P.Gritzman and J. M. Wills, Lattice points, In *Handbook of Convex Geometry*, *Vol.* A, B, eds. P. M. Gruber and J. M. Wills, North-Holland, Amsterdam, 1993, pp. 765–797.
4. J. Hammer, *Unsolved problems concerning lattice points*, Pitman, London, 1977.
5. R. Kannan and L. Lovasz, Covering minima and lattice-point-free convex bodies, *Ann. Math.* **128** (1988) 577–602.
6. H. Minkowski, *Geometrie der Zahlen*, Teubner, Leipzig, 1911.
7. P. R. Scott, A lattice problem in the plane, *Mathematika* **20** (1973) 247–252.
8. P. R. Scott, Sets of constant width and inequalities, *Quart. J. Math. Oxford Ser.* (2) **32** (1981) 345–348.
9. P. R. Scott, Modifying Minkowski's Theorem, *J. Number Theory* **29** (1988) 13–20.
10. M. Sholander, On certain minimum problems in the theory of convex curves, *Trans. Amer. Math Soc.* **73** (1952) 139–173.
11. I. M. Yaglom and V. G. Boltyanskii, *Convex Figures*. Translated by P. J. Kelly and L. F. Walton, Holt, Rinehart and Winston, New York, 1962.
12. M. Yamanouti, Notes on closed convex figures, *Proc. Physico-Math Soc. Japan Ser* (3) **14** (1932) 605–609.

Division of Mathematics, National Institute of Education, 469 Bukit Timah Road, Singapore 259756
awyongpw@nievax.nie.ac.sg

Cutting a Polyomino into Triangles of Equal Areas

Sherman K. Stein

In 1970 Monsky proved that a square cannot be cut into an odd number of triangles of equal areas [**1**], [**6**, p. 118]. This result has been generalized four times. Mead proved that when an n-dimensional cube is cut into simplices of equal volumes, the number of simplices is a multiple of $n!$ [**2**]. Kasimatis proved that when a regular n-gon, $n \geq 5$, is cut into triangles of equal areas, the number of triangles is a multiple of n [**3**]. Stein proved that the theorem about the square

holds for any centrally symmetric polygon with at most eight sides [4] and Monsky generalized this to any centrally symmetric polygon [5]. In this note we extend the theorem about squares to polyominoes consisting of an odd number of squares.

By a *standard square* in the xy-plane we mean a unit square whose corners have integer coordinates. A *standard segment* is a line segment of unit length joining two points with integer coordinates. A *polyomino* is the union of a finite number of standard squares.

Conjecture 1. When a polyomino is cut into triangles of equal areas, the number of triangles is even.

That is a special case of

Conjecture 2. When a polygon in the xy-plane that is bounded by lines parallel to the axes is cut into triangles of equal areas, the number of triangles is even.

The following theorem confirms a special case of the first conjecture.

Theorem. *When a polyomino consisting of an odd number of standard squares is cut into triangles of equal areas, the number of triangles is even.*

We use the machinery described in [1] and [6, pp. 110–117], which we summarize briefly. Define $\varphi : Q \to Q$ by $\varphi(2^a b/c) = a$, where b and c are odd, and $\varphi(0) = \infty$. For instance, $\varphi(2) = 1$, $\varphi(3) = 0$, and $\varphi(5/2) = -1$. Label a point $(x, y) \in Q \times Q$ by P_0 if $\varphi(x) > 0$ and $\varphi(y) > 0$, by P_1 if $\varphi(x) \leq 0$ and $\varphi(y) \geq \varphi(x)$, and by P_2 if $\varphi(x) > \varphi(y)$ and $\varphi(y) \leq 0$. For example, $(2, 0)$ is labeled P_0, $(1, 3)$ is labeled P_1, and $(2, 1)$ and $(1, 1/2)$ are labeled P_2. It can be shown that if a line segment formed of standard segments has ends labeled P_1 and P_2, then the ends of the individual segments are labeled either P_1 or P_2 and an odd number of them have both labels. The following lemma [6, p. 118] is the key tool in establishing the theorem.

Lemma. *Let a polyomino R have area A. Assume that on the boundary of R are an odd number of standard edges with ends labeled P_1 and P_2. Then $\varphi(n) \geq \varphi(2A)$ if R is cut into n triangles of equal areas.*

Proof of the theorem: As may be checked, the only standard segments whose ends are labeled P_1 and P_2 are parallel to the x-axis and lie on lines with an odd y-coordinate. Thus on the border of each standard square is one edge with the labels P_1 and P_2. Edges in the interior of R are adjacent to two standard squares, while edges on the boundary are adjacent to one standard square in R. Since there are an odd number of standard squares in R, the assumption of the lemma holds. Because A is an integer, $\varphi(2A) \geq 1$. Thus the number of triangles is even. ■

The theorem also holds for polyominoes consisting of at most 6 standard squares. (Incidentally, the even case implies the odd one since each standard square can be cut into four congruent squares.) As an illustration, consider the polyomino of area 6 formed of a row of four standard squares with a square attached at each end on the same side of the row, as shown in Figure 1. Note that it can be cut into six triangles of equal areas. No matter how we rotate or translate the polyomino, the hypothesis of the lemma cannot apply, for the conclusion would

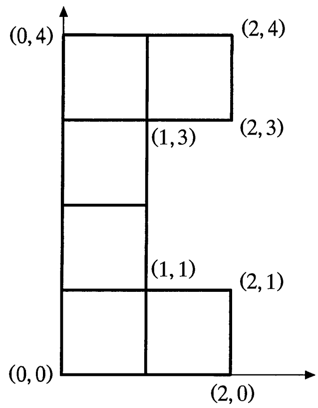


Figure 1

be that the number of triangles would be a multiple of 4. It is necessary to transform the polyomino so that the image has an area A for which $\varphi(A) \leq 0$. Applying the transformation $(x, y) \to (x, y/2)$ followed by the translation by $(1, 1)$ produces a labeling described in the hypothesis of the lemma, as may be verified.

REFERENCES

1. P. Monsky, On dividing a square into triangles, *Amer. Math. Monthly* **77** (1970) 161–164.
2. D. G. Mead, Dissection of a hypercube into simplices, *Proc. Amer. Math. Soc.* **76** (1979) 302–304.
3. E. A. Kasimatis, Dissections of regular polygons into triangles of equal areas, *Discrete and Computational Geometry* **4** (1989) 375–381.
4. S. K. Stein, Equidissections of centrally symmetric octagons, *Aequationes Math.* **37** (1989) 313–318.
5. P. Monsky, A conjecture of Stein on plane dissections, *Math. Zeit.* **205** (1990) 583–592.
6. S. K. Stein and S. Szabó, *Algebra and Tiling*, Math. Assoc. America, 1994.

University of California at Davis, 1 Shields Ave, Davis, CA 95616-8633
stein@math.ucdavis.edu

A Short Proof of Turán's Theorem

William Staton

Extremal graph theory is the search for the thresholds in edge density where substructures of interest are forced to appear in graphs. The canonical extremal theorem involving structure S is of the type: If G is a graph with n vertices containing no S, then G has no more than $f(n)$ edges. The genesis moment of extremal graph theory occurred in 1941 with Turán's article [**1**] in which he proved the canonical extremal result for $S = K_r$, a complete graph with r vertices. The purpose of this note is to provide a new and perhaps shorter proof than has previously been noticed.

Theorem (Turán, 1941). *Graphs with n vertices containing no K_r have no more than $(r-2)n^2/(2r-2)$ edges , for $r \geq 2$.*

Proof: Induct on r. If $r = 2$, the result is obvious. Now if the statement is true for K_r-free graphs it must be shown that K_{r+1}-free graphs have no more than $(r-1)n^2/2r$ edges. Let G be such a graph, and let x be the number of vertices in a largest K_r-free induced subgraph of G. Since the neighbors of any vertex induce a K_r-free subgraph, no vertex of G has degree exceeding x. Let A be a largest induced K_r-free subgraph of G. By induction, there are at most $(r-2)x^2/(2r-2)$ edges in A. Each edge of G not in A is incident with at least one of the $n-x$ vertices not in A, so summing the degrees of these vertices counts each such edge at least once. Hence there are at most $x(n-x)$ such edges and so G has at most $(r-2)x^2/(2r-2) + x(n-x)$ edges. Since

$$\frac{r-2}{2r-2}(x^2) + x(n-x) = \frac{r-1}{2r}n^2 - \frac{r}{2r-2}\left(x - \frac{(r-1)n}{r}\right)^2,$$

the result follows. ■

Turán's theorem continues, in every graph theory textbook, to be the centerpiece of the presentation of extremal graph theory. For this reason, we hope our short proof will be found worthwhile.

REFERENCE

1. P. Turán, On an Extremal Problem in Graph Theory, *Matematicko Fizicki Lapok* **48** (1941) 436–452.

University of Mississippi, University, MS 38677
mmstaton@olemiss.edu

A Characterization of the Set of Points of Continuity of a Real Function

Sung Soo Kim

In this note, we prove the converse of the following well known result: the set of points of continuity of an arbitrary real valued function on a metric space is a countable intersection of open sets [**1**, p. 58].

Lemma. *If X is a nonempty metric space without isolated points, then X has a dense subset A whose complement is also dense in X.*

Proof: Call a set $S \subset X$ an ϵ-net if (a) $d(x, y) \geq \epsilon$ for any two distinct points x, y of S, and (b) S is maximal with respect to (a). Zorn's Lemma yields that ϵ-nets exist for every $\epsilon > 0$. Suppose we have disjoint sets $S_1, S_2, \ldots, S_k$, where each S_i is an $(1/i)$-net. The complement of $S_1 \cup \cdots \cup S_k$ is then nonempty and has no isolated points, and therefore there is an S_{k+1}, disjoint from $S_1 \cup \cdots \cup S_k$, which is an $(1/(k+1))$-net. Then $A = \bigcup_{n=1}^{\infty} S_{2n}$ and $B = \bigcup_{n=1}^{\infty} S_{2n-1}$ are disjoint, and both are dense in X.

Theorem. *Let X be a nonempty metric space without isolated points. If G is a countable intersection of open sets, then there is a function $\phi(x)$ which is continuous exactly on G.*

Proof: Since G is a countable intersection of open sets, we can denote the complement of G as a union of increasing sequence of closed sets F_n, $n = 1, 2, \ldots$. If we define a function $g : X \to R$ by $g(x) = \sum_{n \in K}(\frac{1}{2})^n$, where $K = \{n : x \in F_n\}$, then $g(x)$ converges to 0 as x goes to a point of G. Choose a subset A of X such that A and A^c are both dense, and let $\phi(x) = g(x)(\chi_A(x) - 1/2)$, where $\chi_A(x)$ is a characteristic function on A. Then every neighborhood of an interior point x of G^c contains a point at which the sign of ϕ is different from the sign of $\phi(x)$. If $x \in \partial G \cap G^c$, then $\phi(x) \neq 0$ and every neighborhood of x contains a point y such that $\phi(y) = 0$. Thus ϕ is not continuous at any point of G^c. On the other hand ϕ is continuous at all points of G since $\phi(x) = 0$ for all $x \in G$, and $\phi(x)$ converges to 0 as x goes to a point of G. Therefore, ϕ is continuous exactly on G. ■

Since a function is continuous at any isolated point, we obtain the following.

Corollary. *Let X be a nonempty metric space. If G is an intersection of countable open subsets and contains all of the isolated points of X, then there is a real valued function on X which is continuous exactly on G.*

REFERENCE

1. W. Rudin, *Real and Complex Analysis*, second edition, McGraw-Hill, New York, 1974.

Hanyang University, Ansan, Kyunggi 425-791, Korea
sskim@newton.hanyang.ac.kr

Snow and Ice and Geometry

[my name] is merely a sound. If you look beyond the sound, you will find the body with its circulation, its movement of fluids. Its love of ice, its anger, its longing, its knowledge about space, its weakness, faithlessness and loyalty. Behind these emotions the unnamed forces rise and fade away, parceled-out and disconnected images of memory, nameless sounds. And geometry. Deep inside us is geometry. My teachers at the university asked us over and over what the reality of geometric concepts was. They asked: Where can you find a perfect circle, true symmetry, an absolute parallel when they can't be constructed in this imperfect, external world?

I never answered them, because they wouldn't have understood how self-evident my reply was, or the enormity of its consequences. Geometry exists as an innate phenomenon in our consciousness. In the external world a perfectly formed snow crystal would never exist. But in our consciousness lies the glittering and flawless knowledge of perfect ice.

If you have strength left, you can look further, beyond geometry, deep into the tunnels of light and darkness that exist within each of us, stretching back toward infinity.

There's so much you could do if you had the strength.

Smilla's Sense of Snow, by Peter Høeg, translated by Tiina Nunnally
Dell Publishing, New York, 1994, pp. 318–319

Contributed by Evan J. Romer, Windsor, NY

THE EVOLUTION OF . . .

Edited by **Abe Shenitzer**

Mathematics, York University, North York, Ontario M3J 1P3, Canada

The Literal Calculus of Viète and Descartes

I. G. Bashmakova and G. S. Smirnova

Translated from the Russian by Abe Shenitzer†

1. The contribution of François Viète (1540–1603). Viète tried to create a new science (he called it ars analytica, or analytic art) that would combine the rigor of the geometry of the ancients with the operativeness of algebra. This analytic art was to be powerful enough to leave no problem unsolved: nullum non problema solvere.

Viète set down the foundations of this new science in his *An introduction to the art of analysis* (In artem analyticem isagoge) of 1591.

In this treatise he created a literal calculus. In other words, he introduced the language of formulas into mathematics. Before him, literal notations were restricted to the unknown and its powers. Such notations were first introduced by Diophantus and were somewhat improved by mathematicians of the 15th and 16th centuries.

The first fundamentally new step after Diophantus was taken by Viète, who used literal notations for parameters as well as for the unknown. This enabled him to write equations and identities in general form. It is difficult to overestimate the importance of this step. Mathematical formulas are not just a compact language for recording theorems. After all, theorems can also be stated by means of words; for example, the formula

$$(a+b)^2 = a^2 + 2ab + b^2 \tag{1}$$

can be expressed by means of the phrase "the square of the sum of two quantities is equal to the square of the first quantity, plus the square of the second quantity, plus twice their product." Shorthand also has the virtue of brevity. What counts is that we can carry out operations on formulas in a purely mechanical manner and obtain in this way new formulas and relations. To do this we must observe three rules: 1) the rule of substitution; 2) the rule for removing parentheses; and 3) the rule for reduction of similar terms. For example, from formula (1) one can obtain in a purely mechanical manner, without reasoning, formulas for $(a+b+c)^2$, for $(a+b)^3$, and so on. In other words, literal calculus replaces some reasoning by mechanical computations. In Leibniz' words, literal calculus "relieves the imagination".

†*Translator's note*. The process of creating a literal algebra was begun by Diophantus and was completed by Viète and Descartes. The contribution of Diophantus was described in "The Birth of Literal Algebra", this MONTHLY 106 (1999) 260–271. Viète and Descartes, created the literal calculus we use today.

This article is practically all of Section 4, Chapter V, and Section 1, Chapter VI of [1].

We can hardly imagine mathematics without formulas, without a calculus. But it was such up until Viète's time. The importance of the step taken by Viète is so fundamental that we consider his reasoning in detail.

Viète adopted the basic principle of Greek geometry, according to which only homogeneous magnitudes can be added, subtracted, and can be in a ratio to one another. As he put it: "Homogena homogenei comparare." As a result of this principle, he divides magnitudes into "species": the 1st species consists of "lengths", i.e., of one-dimensional magnitudes. The product of two magnitudes of the 1st species belongs to the 2nd species, which consists of "plane magnitudes", or "squares", and so on.

In modern terms, the domain V of magnitudes considered by Viète can be described as follows:

$$V = \mathbf{R}_+^{(1)} \cup \mathbf{R}_+^{(2)} \cup \cdots \cup \mathbf{R}_+^{(k)} \cup \cdots,$$

where $\mathbf{R}_+^{(k)}$ is the domain of k-dimensional magnitudes, $k \in \mathbf{N}_+$. In each of the domains $\mathbf{R}_+^{(k)}$ we can carry out the operations of addition and of subtraction of a smaller magnitude from a larger one, and can form ratios of magnitudes. If $\alpha \in \mathbf{R}_+^{(k)}$ and $\beta \in \mathbf{R}_+^{(l)}$, then there is a magnitude $\gamma = \alpha\beta$ and $\gamma \in \mathbf{R}_+^{(k+l)}$. If $k > l$, then there exists a magnitude $\delta = \alpha : \beta$, and $\delta \in \mathbf{R}_+^{(k-l)}$.

After constructing this "ladder", Viète proposes to denote unknown magnitudes by vowels A, E, I, O... and known ones by consonants $B, C, D, \ldots$ Furthermore, to the right of the letter denoting a magnitude he places a symbol denoting its species. Thus if $B \in \mathbf{R}_+^{(2)}$, then he writes B plan (i.e., planum–plane), and if an unknown $A \in \mathbf{R}_+^{(2)}$, then he writes A quad (square). Similarly, magnitudes in $\mathbf{R}_+^{(3)}$ get the indices solid or cub and those in $\mathbf{R}_+^{(4)}$ get the indices plano-planum or quadrato-quadratum, and so on.

For addition and subtraction Viète adopts the cossist symbols $+$ and $-$ and introduces the symbol $=$ for the absolute value of the difference of two numbers, thus $B = D$ is the same as $|B - D|$. For multiplication he uses the word "in", A in B, and for division the word "applicare".

Next he introduces the rules

$$B - (C \pm D) = B - C \mp D; \quad B \text{ in } (C \pm D) = B \text{ in } C \pm B \text{ in } D,$$

as well as operations on fractions, written by means of letters, e.g.,

$$\frac{B\text{pl}}{D} + Z = \frac{B\text{pl} + Z \text{ in } D}{D}.$$

Viète's next treatise was *Ad logisticam speciosam notae priores*, which appeared only in 1646 as part of his collected works. In it he set down some of the most important algebraic formulas, such as:

$$(A+B)^n = A^n \pm nA^{n-1}B + \cdots \pm B^n, \quad n = 2,3,4,5;$$
$$A^n + B^n = (A+B)(A^{n-1} - A^{n-2}B + \cdots \pm B^{n-1}), \quad n = 3,5;$$
$$A^n - B^n = (A-B)(A^{n-1} + A^{n-2}B + \cdots + B^{n-1}), \quad n = 2,3,4,5.$$

Viète's literal calculus was perfected by René Descartes (1596–1650), who dispensed with the principle of homogeneity and gave the literal calculus its modern form.

2. The contribution of René Descartes (1596–1650). The 16th century was marked by remarkable achievements in algebra and was followed by a period of relative calm in this area. Most of the energy of 17th-century mathematicians was absorbed by infinitesimal analysis, which was created at that time. Nevertheless, while

inconspicuous at first sight, profound changes were taking place in algebra that can be characterized by one word—arithmetization.

The first steps in this direction were taken by the famous philosopher and mathematician René Descartes (1596–1650). In his *Geometry* (the fourth part of his 1637 *Discourse on method*), whose essential content was the reduction of geometry to algebra or, in other words, the creation of analytic geometry, he first of all transformed Viète's calculus of magnitudes (logistica speciosa). Descartes represented all magnitudes by segments, and constructed a calculus of segments that differed essentially from the one that was used in antiquity and that formed the basis of Viète's construction. Descartes' idea was that the operations on segments should be a faithful replica of (we would say "should be isomorphic to") the operations on rational numbers. Whereas the ancients and Viète regarded the product of two segment magnitudes as an area, i.e., as a magnitude of dimension 2, Descartes stipulated that it was to be a segment. To this end, he introduced a unit segment—which we will denote by e—and defined the product of segments a and b as the segment c that was the fourth proportional to the segments e, a, and b. Specifically (see Figure 1), he constructed an arbitrary angle ABC, and laid off the

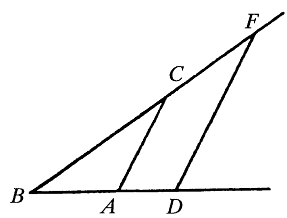


Figure 1

segments $AB = e$, $BD = b$ and $BC = a$. Then he joined A to C, drew $DF \| AC$, and obtained the segment $BF = c = ab$. This meant that the product belonged to the same domain of magnitudes (segments) as the factors. Division was defined analogously: to divide $BF = c$ by $BD = b$ we lay off from the vertex B of the angle the segment $BA = e$, join F to D, and draw AC parallel to DF. The segment BC is the required quotient. In this way, Descartes made the domain of segments into a replica of the semifield $\mathbf{R}^+$. Later he also introduced negative segments (with directions opposite to those of the positive segments) but did not go into the details of operations with negative numbers. Finally, Descartes showed that the operation of extraction of roots (of positive magnitudes) does not take us outside the domain of segments. (We interpolate a comment. Long before Descartes, Bombelli introduced similar rules of operation with segments. Until recently it was thought that he did this in the fourth part of his manuscript published only in the 20th century. However, G. S. Smirnova showed recently that such operations on segments occur also in the parts of Bombelli's *Algebra* published in 1572, i.e., during his lifetime.) To extract the root of $c = BF$, Descartes extended this segment, laid off $FA = e$ on the extension, drew a semicircle with diameter BA, and erected at F the perpendicular to BA. If I is its intersection with the semicircle, then $FI = \sqrt{c}$.

Descartes' calculus was of tremendous significance for the subsequent development of algebra. It not only brought segments closer to numbers but also lent to algebra the simplicity and operativeness that we take advantage of to this day. Another convention introduced by Descartes and used to this day is denoting

unknowns by the last letters of the alphabet: x, y, z, and knowns by the first letters: a, b, c. The only difference between Descartes' symbolism and modern symbolism is his equality sign: $\propto$.

Essentially, it was Descartes who established the isomorphism between the domain of segments and the semifield $\mathbf{R}^+$ of real numbers. However, he gave no general definition of number. This was done by Newton in his *Universal Arithmetic* in which the construction of algebra on the basis of arithmetic reached its completion. He wrote: "Computation is conducted either by means of numbers, as in ordinary arithmetic, or through general variables, as is the habit of analytical mathematicians." And further: "Yet arithmetic is so instrumental to algebra in all its operations that they seem jointly to constitute but a single, complete computing science, and for that reason I shall explain both together."

Newton immediately gives a general definition of number. We recall that in antiquity number denoted a collection of units (i.e., natural numbers), and that ratios of numbers (rational numbers) and ratios of like quantities (real numbers) were not regarded as numbers. Claudius Ptolemy (2nd century AD) and Arab mathematicians did identify ratios with numbers, but in 16th- and 17th-century Europe the Euclidean tradition was still very strong. Newton was the first to break with it openly. He wrote:

> By a 'number' we understand not so much a multitude of units as the abstract ratio of any quantity to another quantity which is considered to be unity. It is threefold: integral, fractional, and surd. An integer is measured by unity, a fraction by a submultiple part of unity, while a surd is incommensurable with unity.

With characteristic brevity, Newton goes on to define negative numbers:

> Quantities are either positive, that is, greater than zero, or negative, that is, less than zero. ... in geometry, if a line drawn with advancing motion in some direction be considered as positive, then its negative is one drawn retreating in the opposite direction.
>
> To denote a negative quantity... the sign $-$ is usually prefixed, to a positive one the sign $+$.

Then Newton formulates rules of operation with relative numbers. We quote his multiplication rule: "A product is positive if both factors are positive or both negative and it is negative otherwise."

He provides no "justifications" for these rules.

(*Translator's note*. The preceding quotations are taken from D. T. Whiteside's English translation of Newton's *Arithmetica universalis*.)

Thus Viète's elaborate domain of magnitudes was replaced in the 17th century by the field of real numbers and arithmetic formed the foundation of algebra.

REFERENCES

1. I. G. Bashmakova and G. S. Smirnova, *The rise and evolution of algebra*. In: *Essays on the history of mathematics*, Moscow State University, Moscow, 1997. (Russian)
2. H. L. L. Busard, *François Viète*. Dictionary of Scientific Biography, Vol. 14, pp. 18 – 25. New York, Scribners, 1970–1990.
3. Theodore M. Brown, *René Descartes*. Dictionary of Scientific Biography, Vol. 4, pp. 51–65. New York, Scribners, 1970–1990.
4. René Descartes, *Discourse on method and meditations*. Translated, with an introduction by L. J. Lafleur. New York, Macmillan, 1986.
5. René Descartes, *Geometry*. Translated by D. E. Smith and M. L. Latham. New York, Dover, 1954.

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**

with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, Richard Holzsager, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Cecil C. Rousseau, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before August 31, 1999; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10718. *Proposed by David M. Bloom, Brooklyn College of CUNY, Brooklyn, NY.* Let p be a prime number with $p \equiv 7 \pmod 8$, and let $L_p = \{1, 2, 3, \dots, (p-1)/2\}$. Prove that the sum of the quadratic residues modulo p in L_p equals the sum of the quadratic nonresidues modulo p in L_p. For example, the quadratic residues in L_{23} are 1, 2, 3, 4, 6, 8, and 9, and the quadratic nonresidues in L_{23} are 5, 7, 10, and 11. Both lists sum to 33.

10719. *Proposed by Jean Anglesio, Garches, France.* Let A, I, and G be three points in the plane. Let M denote the point 2/3 of the way from A to I, and let U and V be the circles of radius $|AM|$ each of which is tangent to AI at M. Show that when G is outside both U and V, there are precisely two triangles ABC with incenter I and centroid G. Provide a Euclidean construction for them. Show that when G is in the interior of U or V, there does not exist a triangle ABC with incenter I and centroid G.

10720. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.* A "binary maze" is a directed graph in which exactly two arcs lead from each vertex, one labeled 0 and one labeled 1. If $b_1, b_2, \dots, b_m$ is any sequence of 0s and 1s and v is any vertex, let $vb_1b_2\cdots b_m$ be the vertex reached beginning at v and traversing arcs labeled $b_1, b_2, \dots, b_m$ in order. A sequence $b_1, b_2, \dots, b_m$ of 0s and 1s is a *universal exploration sequence* of order n if, for every strongly connected binary maze on n vertices and every vertex v, the sequence

$$v, vb_1, vb_1b_2, \dots, vb_1b_2\cdots b_m$$

includes every vertex of the maze. For example, 01 is a universal exploration sequence of order 2, and it can be shown that 0110100 is universal of order 3.

(a) Prove that universal exploration sequences of all orders exist.

(b)* Find a good estimate for the asymptotic length of the shortest such sequence of order n.

10721. *Proposed by Daniel A. Sidney, Massachusetts Institute of Technology, Cambridge, MA.* Let $f(x) = \sin x/x$, and let m and n be nonnegative integers. Compute

$$\int_0^\infty \frac{d^m}{dx^m} f(x)\, \frac{d^n}{dx^n} f(x)\, dx.$$

10722. *Proposed by Richard F. McCoart, Loyola College, Baltimore, MD.*
(**a**) In how many ways can $2n$ indistinguishable balls be placed into n distinguishable urns, if the first r urns may contain at most $2r$ balls for each $r \in \{1, 2, \ldots, n\}$?
(**b**) Suppose that $0 \leq m \leq n$. In how many of the ways enumerated in part (**a**) are exactly m urns empty?

10723. *Proposed by Christopher J. Hillar, Yale University, New Haven, CT.* Let p be an odd prime. Prove that $\sum_{i=1}^{p-1} 2^i \cdot i^{p-2} \equiv \sum_{i=1}^{(p-1)/2} i^{p-2} \pmod{p}$.

10724. *Proposed by Serge Tabachnikov, University of Arkansas, Fayetteville, AR.*
(**a**) Let P be a convex plane polygon with vertices $A_1, \ldots, A_n$, and let l be a continuous transverse field of directions along the boundary ∂P. (This means that through every point $X \in \partial P$ there passes a line $l(X)$ that intersects the interior of P and depends continuously on X.) Let α_i and β_i be the angles between the line $l(A_i)$ and the adjacent sides A_iA_{i-1} and A_iA_{i+1}, respectively. Assume that $\prod_1^n \sin\alpha_i = \prod_1^n \sin\beta_i$. Prove that the lines $l(X)$ cover the interior of P twice, that is, every interior point of P belongs to at least two of these lines.
(**b**) Suppose $n \geq 3$, and let P be a convex polyhedron in n-dimensional space. As in (**a**), a continuous transverse line field l is given along the boundary ∂P. This field has the property that for every $(n-2)$-dimensional face E of P there exists a hyperplane $\pi(E)$ such that all the lines $l(X)$ with $X \in E$ belong to $\pi(E)$. Prove that the lines $l(X)$ cover the interior of P twice.

SOLUTIONS

Principal Ideals in Noetherian Rings

10534 [1996, 510]. *Proposed by Paul Arne Østvær, Oslo University, Oslo, Norway.* Suppose that R is a Noetherian ring in which all maximal ideals are principal. Show that all ideals in R are principal.

Solution by Robert Gilmer, Florida State University, Tallahassee, FL. If $M = (m)$ is a maximal ideal of R, then M/M^2 is a vector space over the field R/M of dimension at most 1. Hence there are no ideals of R properly between M and M^2. From this it follows (R. Gilmer, *Multiplicative Ideal Theory,* Queen's Papers Pure Appl. Math. **90** (1992), Theorem 39.2) that $R = D_1 \oplus \cdots \oplus D_n \oplus S_1 \oplus \cdots \oplus S_m$ is a finite direct sum of Dedekind domains D_i and special primary rings S_i. To show that each ideal of R is principal, it suffices to show that the D_i and S_i have this property. For S_i this is part of the definition of a special primary ring (Gilmer, p. 200). Moreover, D_i inherits from R the property that each of its maximal ideals is principal, and a Dedekind domain is a principal ideal domain whenever all of its maximal ideals are principal.

Editorial comment. D. D. Anderson mentions a stronger result that appears in R. Gilmer and W. Heinzer, Principal ideal rings and a condition of Kummer, *J. Algebra* **83** (1983) 285–292: If R has the ascending chain condition on *principal* ideals and each maximal ideal of R is principal, then every ideal of R is principal.

Solved also by Mahalal'el ben keinan (Israel), F. Calegari (Australia), J. E. Dawson (Australia), T. H. Foregger, O. Moubinool (France), S. Sertöz (Turkey), and M. Tabaâ (Morocco).

A Telescoping Constraint

10566 [1997, 68]. *Proposed by Gerry Myerson, Macquarie University, Australia.* Let S be a finite set of cardinality $n > 1$. Let f be a real-valued function on the power set of S, and suppose that $f(A \cap B) = \min\{f(A), f(B)\}$ for all subsets A and B of S. Prove that

$$\sum (-1)^{n-|A|} f(A) = f(S) - \max f(A),$$

where the sum is taken over all subsets A of S and the maximum is taken over all proper subsets A of S.

Composite solution by Reiner Martin, Deutsche Bank, London, U. K., and Walter Stromquist, Berwyn, PA. Since $f(A)$ cannot increase when A decreases, the maximum over proper subsets occurs at some set U of size $n-1$; let x be the missing element. For $B \subseteq U$, let $B' = B \cup \{x\}$. Since $B = B' \cap U$, we have $f(B) = \min\{f(B'), f(U)\} = f(B')$ unless $B = U$. Since B and B' differ in size by 1, the terms of the summation other than for S and U cancel in pairs, and we have $\sum(-1)^{n-|A|} f(A) = f(S) - f(U) = f(S) - \max f(A)$.

Solved also by R. Barbara (France), D. Beckwith, M. Benedicty, J. C. Binz (Switzerland), M. Bowron, M. A. Brodie, D. Callan, R. J. Chapman (U. K.), J. E. Dawson (Australia), D. Donini (Italy), D. Dwyer, R. Ehrenborg, G. Gordon, R. Holzsager, T. Jager, W. Janous, K. S. Kedlaya, N. Komanda, O. Kouba (Syria), J. Kuplinsky, C. Lanski, J. H. Lindsey II, O. P. Lossers (The Netherlands), A. Nijenhuis, K. Schilling, Z. Shan & E. Wang (Canada), M. Shemesh (Israel), J. Simpson (Australia), J. H. Steelman, H.-T. Wee (Singapore), GCHQ Problems Group (U. K.), WMC Problems Group, and the proposer.

A Summation Identity

10575 [1997, 169]. *Proposed by Xiaokang Yu, Penn State Altoona Campus, Altoona, PA.* Prove that

$$\sum_{l=0}^{n}(-1)^l\binom{n}{l}(2l)!\sum_{m=0}^{2l}\frac{(-1)^m}{m!} = \sum_{l=0}^{n}(-1)^l\binom{n}{l}2^{n-l}(n+l)!$$

for every nonnegative integer n.

Solution I by Richard Holzsager, American University, Washington, DC. The expression $(2l)!\sum_{m=0}^{2l}(-1)^m/m!$ on the left side is the well-known *derangement number* D_{2l}, the number of ways to return hats to $2l$ people so that no hat returns to its owner. When hats are returned to n couples, let $d(n,l)$ be the number of permutations in which the hats for the people in l couples are deranged, while the remaining hats all go to their owners. The left side is $\sum_{l=0}^{n}(-1)^l d(n,l)$, counting such permutations positively or negatively depending on the parity of the number of deranged couples.

The expression $\binom{n}{l}2^{n-l}(n+l)!$ on the right side is the number of ways to permute the $2n$ hats so that one specified person in each of $n-l$ couples receives the right hat (others may also receive the right hat). In this sum, these terms are weighted by the parity of the number of couples with no member specified. In order to prove the desired identity, it suffices to prove that, in the sum on the right, each permutation of the $2n$ hats is counted with total weight 1, -1, or 0, depending on whether the set of those who get the wrong hat forms an even number of couples, an odd number of couples, or includes some "isolated" individuals whose mates get the right hat.

Consider a permutation of hats in which the correct hats are received by exactly i couples and j isolated individuals. This permutation is counted for each of the $2^r\binom{i}{r}\binom{j}{s}$ choices of one partner from each of r of the couples and s of the individuals; such a choice contributes weight $(-1)^{n-(r+s)}$. The total weight for this permutation is therefore

$$\begin{aligned}\sum_{r\le i}\sum_{s\le j}(-1)^{n-(r+s)}2^r\binom{i}{r}\binom{j}{s} &= (-1)^n\sum_{r\le i}(-2)^r\binom{i}{r}\sum_{s\le j}(-1)^s\binom{j}{s} \\ &= (-1)^n(1-2)^i(1-1)^j = (-1)^{n-i}0^j,\end{aligned}$$

with $0^0 = 1$. This gives ± 1 or 0 in precisely the cases we need.

Solution II by M. A. Prasad, Mumbai, India. We first note that

$$\frac{(2l)!}{m!} = \binom{2l}{m}(2l-m)! = \binom{2l}{m}\int_0^\infty e^{-t}t^{2l-m}\,dt,$$

using the integral representation of the gamma function. Thus, we can rewrite the left side L of the desired equation as

$$L=\sum_{l=0}^{n}(-1)^l\binom{n}{l}\sum_{m=0}^{2l}(-1)^m\binom{2l}{m}\int_0^\infty e^{-t}t^{2l-m}\,dt.$$

Interchanging the order of integration and summation and summing over m by the Binomial Theorem yields

$$L=\int_0^\infty e^{-t}\sum_{l=0}^{n}(-1)^l\binom{n}{l}\sum_{m=0}^{2l}(-1)^m\binom{2l}{m}t^{2l-m}\,dt=\int_0^\infty e^{-t}\sum_{l=0}^{n}(-1)^l\binom{n}{l}(1-t)^{2l}\,dt.$$

By the Binomial Theorem, the inner sum is $\left(1-(1-t)^2\right)^n$, which simplifies to $t^n(2-t)^n$. Expanding $(2-t)^n$ and then interchanging the order of summation and integration yields

$$L=\int_0^\infty e^{-t}t^n\sum_{l=0}^{n}\binom{n}{l}(-1)^l2^{n-l}t^l\,dt=\sum_{l=0}^{n}(-1)^l\binom{n}{l}2^{n-l}\int_0^\infty e^{-t}t^{n+l}\,dt.$$

Since the remaining integral is the integral formula for $(n+l)!$, we have the desired result.

Solution III by Doron Zeilberger, Temple University, Philadelphia, PA. We use the umbral calculus (the calculus of finite differences). Define the shift operator E by $Ef(x)=f(x+1)$. Since E commutes with multiplication by constants, we can apply the binomial theorem to expressions involving E and operators that multiply by constants.

Let $\Delta=1-E$. Working solely with operators, we have

$$\begin{aligned}R:=\sum(-1)^l\binom{n}{l}2^{n-l}E^{n+l}&=2^nE^n(1-\frac{1}{2}E)^n=E^n(2-E)^n\\&=(1-\Delta)^n(1+\Delta)^n=(1-\Delta^2)^n=\sum(-1)^l\binom{n}{l}\Delta^{2l}.\end{aligned}$$

Let F denote the factorial function. The right side of the desired equation is RF evaluated at 0. To complete the proof, it suffices to show that $\Delta^{2l}F$ evaluated at 0 yields $(2l)!\sum_{m=0}^{2l}(-1)^m/m!$. Since $\Delta^{2l}=(E-1)^{2l}=\sum_{m=0}^{2l}(-1)^m\binom{2l}{m}E^{2l-m}$, we have

$$\Delta^{2l}F(0)=(2l)!\sum_{m=0}^{2l}\frac{(-1)^m}{m!}\frac{1}{(2l-m)!}E^{2l-m}F(0)=(2l)!\sum_{m=0}^{2l}\frac{(-1)^m}{m!}.$$

Solution IV (composite) by Yong Kong, Washington University, St. Louis, MO, and John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. We use an auxiliary identity

$$\sum_{l=0}^{n}\binom{n}{l}\binom{2l}{m}(-1)^l=(-1)^n2^{2n-m}\binom{n}{m-n}.$$

The two sides of this identity agree for all n when $m\in\{0,1\}$, and both satisfy the recurrence $f(n,m)=-2f(n-1,m-1)-f(n-1,m-2)$. Thus they are always equal.

To prove the desired identity, we replace m by $2l-m$ on the left side, then interchange the order of summation, and apply the auxiliary identity as follows:

$$\begin{aligned}\sum_{l=0}^{n}(-1)^l\binom{n}{l}(2l)!\sum_{m=0}^{2l}\frac{(-1)^m}{m!}&=\sum_{l=0}^{n}(-1)^l\binom{n}{l}\sum_{m=0}^{2l}\binom{2l}{m}m!(-1)^{2l-m}\\&=\sum_m(-1)^mm!\sum_l\binom{n}{l}\binom{2l}{m}(-1)^l=\sum_m(-1)^mm!(-1)^n2^{2n-m}\binom{n}{m-n}\end{aligned}$$

Finally, in the last expression set $l = m - n$.

Editorial comment. William Seaman and the proposer proved that both sides equal the value at $x = -1$ of $\sum_{m=0}^{2n} \left(\frac{d}{dx}\right)^m \left(1 - x^2\right)^n$.

Solved also by J. C. Binz (Switzerland), R. J. Chapman (U. K.), Q. H. Darwish (Oman), J. E. Dawson (Australia), M. Ismail & P. Simeonov (U. K.), M. Omarjee (France), L. Pebody (U. K.), C. R. Pranesachar (India), R. Richberg (Germany), W. J. Seaman, H.-J. Seiffert (Germany), A. Tissier (France), and the proposer.

A Large Bipartite Subgraph

10580 [1997, 270]. *Proposed by Stephen C. Locke, Florida Atlantic University, Boca Raton, FL.* Let G be a simple graph with v vertices and e edges and with maximum degree at most 3. Suppose that no component of G is a complete graph on 4 vertices. Prove that G contains a bipartite subgraph with at least $e - v/3$ edges.

Solution by James M. Benedict and Gerald Thompson, Augusta State University, Augusta, GA. When G is bipartite, the claim holds trivially, so we may assume that the chromatic number of G is at least 3. Since G does not have a complete graph of order 4 as a component, Brooks's Theorem implies that G is 3-colorable. Consider a proper 3-coloring using colors red, white, and blue; we may assume that blue appears least often.

Each blue vertex has at most 3 neighbors, all red or white. In either red or white it has at most one neighbor. After removing that edge, we can change the blue vertex to that color and still have a proper coloring. Doing this for each blue vertex deletes at most $v/3$ edges and produces a 2-colored (that is, bipartite) subgraph.

Editorial comment. Brooks's Theorem states that a graph with maximum degree k has a proper k-coloring if $k \geq 3$ and no component is a complete graph of order $k + 1$ (see for example J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, North-Holland, 1976, p. 122). An inductive solution that avoids Brooks's Theorem is also possible.

Solved also by R. J. Chapman (U. K.), C. P. Rupert, P. Tracy, and the proposer.

Solid Angles of a Tetrahedron

10598* [1997, 457]. *Proposed by Jeffrey C. Lagarias, AT&T Research, and Thomas J. Richardson, Bell Laboratories, Murray Hill, NJ.* Let F_1, F_2, F_3, F_4 denote the faces of a tetrahedron. For $i = 1, 2, 3, 4$, let α_i denote the solid angle of the vertex opposite face F_i, where the measure of a solid angle is normalized so that a full solid angle is 1, and let β_i denote the area of F_i, where the unit of area is normalized so that the tetrahedron has surface area 1.

(a) Prove that $\beta_i \geq \alpha_i$.

(b) Generalize to m dimensions.

Solution by John H. Lindsey II, Ft. Myers, FL.

(a) We prove the sharper claim that $\beta_i > f(\pi\alpha_i)$, where $f(\theta) = \sec\theta\tan\theta - \tan^2\theta = 1/(\csc\theta + 1)$. To see that this bound is sharper, note that $\alpha_i < 1/2$, since $1/2$ is the normalized solid angle of a plane and each angle of the tetrahedron lies on one side of a plane. Since $f(0) = 0$, $f(\pi/2) = 1/2$, and $f''(\theta) = \sec^4\theta(\sin\theta - 1)^2(\sin\theta - 2) < 0$, we have $f(\pi\alpha) \geq \alpha$ for $0 < \alpha < 1/2$.

Suppose that a counterexample exists. We relabel and translate to arrange that the counterexample occurs for $i = 1$, the vertex opposite F_1 is the origin O, and the other vertices are xU, yV, zW, where U, V, W are unit vectors and x, y, z are positive. Then

$$\frac{1}{1/\beta_1 - 1} = \frac{\beta_1}{\beta_2 + \beta_3 + \beta_4} = \frac{|(xU - zW) \times (yV - zW)|}{xy|U \times V| + xz|U \times W| + yz|W \times V|}. \quad (1)$$

Varying x, y, and z does not change α_1; therefore we may choose a sequence of counterexamples $(O, x_n U, y_n V, z_n W)$ with $0 < x_n, y_n, z_n$ for which β_1 converges to its infimum. Some ordering of (x_n, y_n, z_n) must occur infinitely often, so after reordering the vertices, passing to a subsequence, and rescaling, we may assume $1 = x_n \geq y_n \geq z_n$.

Suppose $z_n \to 0$. Then terms that are small relative to $x_n y_n |U \times V|$ do not affect the limit of (1). Ignoring them, we are left with

$$\frac{|x_n U \times (y_n V - z_n W)|}{x_n y_n |U \times V| + x_n z_n |U \times W|}.$$

This is a 2-dimensional analogue $((1/\beta'_{1,n}) - 1)^{-1}$ for the triangle with edges $x_n y_n U \times V$ and $x_n z_n U \times W$. Assuming the 2-dimensional version, we have

$$\lim \frac{1}{\frac{1}{\beta_{1,n}} - 1} = \lim \frac{1}{\frac{1}{\beta'_{1,n}} - 1} \geq \lim \frac{1}{\frac{1}{f(\pi\alpha'_{1,n})} - 1} = \frac{1}{\frac{1}{f(\pi\alpha'_1)} - 1} > \frac{1}{\frac{1}{f(\pi\alpha_1)} - 1},$$

since α'_1, the angle between $U \times V$ and $U \times W$, is one of the dihedral angles of the tetrahedron that meet at O. Thus we do not have a counterexample. Therefore we may assume that z_n does not converge to zero, and passing to a subsequence we get a limiting nondegenerate tetrahedron that is a counterexample and that is a minimum for β_1.

Let P_1, P_2, P_3, P_4 be the vertices of this tetrahedron, with P_1 the origin. We may assume that F_1 is parallel to the x,y-plane and at distance a from it. Let P_1^* be the projection of P_1 onto the plane containing F_1. Imagine moving P_i toward P_1 at a constant rate. By minimality, the derivative of the ratio of areas defining $(1/\beta_1 - 1)^{-1}$ with respect to time is 0 at the start of this motion. The component of the movement orthogonal to the plane containing F_1 has no first order effect on the area of F_1 at the start, so if we replace the area of F_1 by its projection on the original plane of F_1, then the derivative of our new ratio is again 0 at the start. However, the new ratio is a quotient of linear functions of time, so, since it has derivative 0 at the start, it must be constant. If P_1^* lies outside F_1, say outside the edge $P_3 P_4$, then while P_2 is en route to P_1, P_2^* (the projection of P_2 onto the plane containing F_1) crosses the extended edge, at which point our new ratio is 0. This is impossible since the ratio is constant. When P_i reaches P_1, our new ratio is the ratio of the area of the projection of F_i onto the plane of F_1 to the area of F_i. This ratio is $b_i(a^2 + b_i^2)^{-1/2}$, where b_i is the distance from P_1^* to the edge of F_1 opposite P_i. It follows that $(1/\beta_1 - 1)^{-1} = b_i(a^2 + b_i^2)^{-1/2}$ for every $i \in \{2, 3, 4\}$. Thus P_1^* and $b = b_i$ are the incenter and inradius of F_1.

Let $g(\gamma)$ be the solid angle, normalized so that the full solid angle is 4π, from P_1 spanned by a right triangle $P_1^* RS$ in the plane of F_1, with $|P_1^* R| = b$, $\angle P_1^* RS = \pi/2$, and $\angle RP_1^* S = \gamma$. A calculation shows that

$$g(\gamma) = \gamma - \int_0^\gamma \frac{a\, d\theta}{\sqrt{a^2 + b^2 \sec^2\theta}} = \gamma - \arcsin\left(\frac{a \sin\gamma}{\sqrt{a^2 + b^2}}\right).$$

Since $g(\gamma)$ is concave upward, $g(0) = 0$, and $g(\pi/2) = \arctan(b/a)$, it follows that $g(\gamma) < (2\gamma/\pi)\arctan(b/a)$ for $\gamma \in (0, \pi/2)$. Since F_1 is a union of six such triangles $P_1^* RS$, with angles γ summing to 2π, we see that $4\pi\alpha_1 = \sum g(\gamma) < (2\sum\gamma/\pi)\arctan(b/a) = 4\arctan(b/a)$, where the summations are taken over the 6 values of γ. Hence $\tan\pi\alpha_1 < b/a$, and

$$\beta_1 = \frac{b}{b + \sqrt{a^2 + b^2}} > \frac{1}{\csc\pi\alpha_1 + 1} = f(\pi\alpha_1).$$

Thus a counterexample cannot exist.

(b) The argument is similar. In dimension $m > 3$ we again get strict inequality. To see this, consider a counterexample in dimension m. Arrange that $\beta_1 \leq f(\pi\alpha_1)$ and that the vertices are $P_1 = O$ and $P_i = z_i U_i$ for $2 \leq i \leq m+1$, where U_i are unit vectors and $z_i > 0$. Again

varying the z_i does not change α_1, so we may choose a sequence for which β_1 approaches its infimum. A subsequence either degenerates to a lower dimensional simplex or leads to a counterexample with β_1 minimal. If the limit is degenerate, then a computation shows that there is a counterexample for lower m, contradicting the minimality of m.

Therefore we may consider a counterexample that has β_1 minimal under varying the z_i. Assume that F_1 lies in the affine subspace $S = \{(a, x_2, \dots, x_m)\}$, and let P_1^* be the projection of P_1 into this subspace. Arguing as for the 3-dimensional case, we see that P_1^* is in the interior of F_1 and is equidistant from all the faces of F_1. Let this common distance be b. Let F be a face of F_1, let T be the $(m-2)$-dimensional affine subspace containing F, and let Q be the orthogonal projection of P_1 into T. Let $f(r)dr$ be the solid angle generated from P_1^* by the points of T whose distance from Q is between r and $r+dr$. Let $S(r)$ be the sphere of radius r about Q in T, and define $g_F(r) = \text{area}(F \cap S(r))/\text{area}(S(r))$. Note that $g_F(r)$ is nonincreasing, by convexity of F. If a solid angle Φ in S with vertex P_1^* meets T at a distance from Q of between r and $r+dr$, then let $h_b(r)$ be the measure of the solid angle from P_1 generated by the portion of Φ bounded by T. With these definitions, F generates a solid angle of $\int_0^\infty g_F(r)f(r)dr$ from P_1^* in S, and the portion of F_1 between F and P_1^* generates a solid angle of $\int_0^\infty g_F(r)h_b(r)f(r)dr$ from P_1.

Since $g_F(r)$ is nonincreasing and nonconstant, and since $h_b(r)$ is increasing, we have

$$\int_0^\infty f(r)dr \int_0^\infty g_F(r)h_b(r)f(r)dr < \int_0^\infty h_b(r)f(r)dr \int_0^\infty g_F(r)f(r)dr.$$

Let A_t be the $(t-1)$-dimensional area of the t-dimensional sphere of radius 1. Summing the last inequality over all faces of F_1 gives

$$A_m\alpha_1 \int_0^\infty f(r)dr < A_{m-1} \int_0^\infty h_b(r)f(r)dr. \tag{2}$$

The same calculation applies if F_1 is replaced by the slab $G = \{(a, x_2, \dots, x_m) : |x_2| \le b\}$, except that (1) we now get equality, since for both faces H of G, the function g_H is identically 1, and (2) α_1 is replaced by ϕ, the probability that the ray from O through a random point $v = (y_1, \dots, y_m)$ on the unit sphere hits G. Hence

$$A_m\phi \int_0^\infty f(r)dr = A_{m-1} \int_0^\infty h_b(r)f(r)dr. \tag{3}$$

From (2) and (3), we infer that $\alpha_1 < \phi$. The random ray from O hits G if and only if $y_1 > 0$ and $|y_2|/y_1 \le b/a$. This depends only on the direction of (y_1, y_2), which is uniformly distributed. Thus $\alpha_1 < \phi = \alpha_1'$, where α_1' is the value of α_1 for the (2-dimensional) isosceles triangle J with altitude a and base $2b$. Since J has the same value of β_1 as G has, we are reduced to the 2-dimensional case. Since we can approximate G by F_1, $f(\pi\alpha_1)$ is the best possible lower bound for β_1.

A Tricky Convergence

10614 [1997, 767]. *Proposed by Grigore-Raul Tataru, University of Bucharest, Bucharest, Romania.* Fix $p > 1$. Suppose that $a_1, a_2, \dots$ is a sequence of positive real numbers such that $a_n a_{n+1} a_{n+2}^p + a_{n+2} - a_n = 0$ for all $n \ge 1$. Show that $\{a_n\}$ is convergent.

Solution by the GCHQ Problems Group, Cheltenham, U. K. Since $a_n - a_{n+2} = a_n a_{n+1} a_{n+2}^p$ is positive, the even and odd subsequences are decreasing and therefore convergent, say to x and y respectively. Taking limits gives $yx^{p+1} = 0 = xy^{p+1}$, so at least one of x and y must be 0. Without loss of generality, we may assume $x = 0$. If $y > 0$, then $a_{2n-1} - a_{2n+1} > y^{p+1}a_{2n}$, so the series $\sum a_{2n}$ converges. Let m be large enough that $a_{2m+1} < 2y$ and $a_{2m} < 1$. Let $\epsilon = a_{2m}$. For $n \ge m$, we have $a_{2n} - a_{2n+2} < 2y\epsilon^{p+1}$, so the number of integers n with $\epsilon/2 \le a_{2n} < \epsilon$ is at least $(\epsilon/2)/(2y\epsilon^{p+1}) = 1/(4y\epsilon^p) > 1/(4y\epsilon)$, and

the sum of these terms is therefore at least $1/(8y)$. Thus there are infinitely many disjoint blocks of terms, each of which sums to at least $1/(8y)$. This contradicts the fact that $\sum a_{2n}$ converges. Therefore $x = y = 0$, and $a_n \to 0$.

Solved also by J. Anglesio (France), S. S. Kim (Korea), K.-W. Lau (Hong Kong), J. H. Lindsey II, M. Shemesh (Israel), NSA Problems Group, and the proposer.

Tails of an Alternating Series

10624 [1997, 871]. *Proposed by William F. Trench, Trinity University, San Antonio, TX.* Suppose that $a_0 > a_1 > a_2 > \cdots$ and $\lim_{n\to\infty} a_n = 0$. Define $S_n = \sum_{j=n}^{\infty}(-1)^{j-n}a_j = a_n - a_{n+1} + a_{n+2} - \cdots$. Show that $\sum a_n S_n < \infty$ if and only if $\sum a_n^2 < \infty$.

Solution by Douglas B. Tyler, Hughes Aircraft Company, El Segundo, CA. More generally, we prove that the three series $\sum S_n^2$, $\sum a_n S_n$, and $\sum a_n^2$ converge or diverge together. By the alternating series test, S_n exists and satisfies $0 < S_n < a_n$. Thus $\sum S_n^2 < \sum a_n S_n < \sum a_n^2$. So it suffices to show that finiteness of $\sum S_n^2$ implies finiteness of $\sum a_n^2$. To prove it, use $S_n = a_n - S_{n+1}$ and the inequality $(x+y)^2 \le 2(x^2+y^2)$ to infer

$$\sum a_n^2 = \sum (S_n + S_{n+1})^2 \le \sum 2(S_n^2 + S_{n+1}^2) = 2\sum S_n^2 + 2\sum S_{n+1}^2 < 4\sum S_n^2.$$

Solved also by S. A. Ali, K. F. Andersen (Canada), R. Barbara (France), G. L. Brody (U. K.), P. Bracken (Canada), P. Budney, D. Callan, R. J. Chapman (U. K.), J. E. Dawson (Australia), M. N. Deshpande (India), Z. Franco, T. Goebeler & T. Siemers, T. Hermann, V. Hernandez & J. Martin (Spain), S. S. Kim (Korea), R. A. Kopas, O. Kouba (Syria), M. Kumar (India), J. H. Lindsey II, N. Lord (U. K.), P. Mengert, R. Mortini (France), M. Omarjee (France), L. Opperman, G. Peng, H. Salle (The Netherlands), K. Schilling, H.-J. Seiffert (Germany), N. C. Singer, A. Stenger, S. J. Swiniarski, N. S. Thornber, A. Tissier (France), P. Trojovsky (Czech Republic), M. Woltermann, C. Xiong, GCHQ Problems Group (U. K.), NSA Problems Group, WMC Problems Group, and the proposer.

Cosecants and Near-Integers

10630 [1997, 975]. *Proposed by Richard Stong, Rice University, Houston, TX.* It is possible to show that $\csc(3\pi/29) - \csc(10\pi/29) = 1.999989433\ldots$. Prove that there are no integers j, k, n with n odd satisfying $\csc(j\pi/n) - \csc(k\pi/n) = 2$.

Solution by Allen Stenger, Tustin, CA. Suppose that there exist such integers j, k, n. Let $\omega_m = e^{2\pi i/m}$ be a primitive mth root of unity. Then

$$\frac{2i}{\omega_{2n}^{j} - \omega_{2n}^{-j}} - \frac{2i}{\omega_{2n}^{k} - \omega_{2n}^{-k}} = 2,$$

which rearranges to

$$i = \frac{(\omega_{2n}^{k} - \omega_{2n}^{-k})(\omega_{2n}^{j} - \omega_{2n}^{-j})}{(\omega_{2n}^{k} - \omega_{2n}^{-k}) - (\omega_{2n}^{j} - \omega_{2n}^{-j})}.$$

This implies that i is in $\mathbb{Q}(\omega_{2n})$, the cyclotomic field of order $2n$ over the rationals.

We now claim that adjoining i to $\mathbb{Q}(\omega_{2n})$ produces $\mathbb{Q}(\omega_{2n}, i) = \mathbb{Q}(\omega_{4n})$. To show this, observe that $\omega_{2n} = \omega_{4n}^2 \in \mathbb{Q}(\omega_{4n})$, $i = \omega_{4n}^n \in \mathbb{Q}(\omega_{4n})$, and $\omega_{4n} = \omega_{4n}^{n+1}/\omega_{4n}^n = \omega_{2n}^{(n+1)/2}/i \in \mathbb{Q}(\omega_{2n}, i)$. (This is where we need the hypothesis that n is odd.)

This produces the contradiction: If it were true that $i \in \mathbb{Q}(\omega_{2n})$, then we would have $\mathbb{Q}(\omega_{2n}) = \mathbb{Q}(\omega_{2n}, i) = \mathbb{Q}(\omega_{4n})$, but this is impossible because the degrees of $\mathbb{Q}(\omega_{2n})$ and $\mathbb{Q}(\omega_{4n})$ over $\mathbb{Q}$ are known to be $\phi(2n)$ and $\phi(4n) = 2\phi(2n)$, respectively.

Editorial comment. Gerry Myerson proved more: If j, k, n are positive integers with no common factor and $\csc(j\pi/n) - \csc(k\pi/n)$ is a nonzero rational number r, then n is 2 or 6 and r is ±1, ±2, ±3, or ±4.

Solved also by R. J. Chapman (U. K.), J. H. Lindsey II, G. Myerson (Australia), and the proposer.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843-3368

An Accompaniment to Higher Mathematics. By George R. Exner. Springer-Verlag, 1996, xvii + 198 pp., $29.95.
Journey into Mathematics: An Introduction to Proofs. By Joseph Rotman. Prentice Hall, 1998, xiii + 237 pp., $66.67.
Mathematical Thinking: Problem Solving and Proofs. By John P. D'Angelo and Douglas B. West. Prentice Hall, 1997, xviii + 365 pp., $71.

Reviewed by Joseph H. Silverman

One of the highest hurdles facing erstwhile mathematics majors is the transitory leap from rote problem solving to proof creation. (Dissenting readers may substitute "discovery" for "creation" if they wish, but the fascinating question of whether proofs are discovered or created is largely irrelevant to our discussion.) By rote problem solving I mean, of course, the sort of process used in most calculus classes whereby students are shown standard problem templates and, after absorbing a sufficient number of examples, learn to solve similar problems in a color-by-number fashion. There is nothing inherently wrong with this activity, since for most students achieving competence requires hard work and a significant amount of mental effort, two essential ingredients of every good college education. Further, students who do master the material are left with a sense of accomplishment, and we hope that those students who actually need the calculus for their further studies in engineering, economics, or the hard sciences will solidify and deepen their understanding of the subject when they see it used in their other courses.

For mathematicians, problem solving of the sort just described is related to mathematics much as doing a crossword puzzle is related to writing a novel. Both activities require a good vocabulary and some mental agility, but only the latter requires creativity. However, just as there are writing guidelines and exercises designed to assist budding novelists, it is the thesis of these three books that there are teachable techniques through which aspiring mathematics majors can make the transition from mathematical watchers to mathematical doers.

What, roughly, are some of the meta-mathematical tools (as opposed to mathematical techniques such as induction) that every mathematician keeps close at hand when tackling a mathematical problem? In no particular order, the following (non-definitive and non-disjoint) list comes to mind:

- Do lots of examples, numerical or otherwise.
- Specialize the problem. Do special cases.
- Generalize the problem. Eliminate unnecessary hypotheses. This technique can be surprisingly effective, since with fewer hypotheses, there are fewer ways to proceed!

- Search for counterexamples to the original problem.
- Find counterexamples when each of the hypotheses is relaxed. Thus the origin of the phrase "the exception 'proves' the rule," using the original sense of the word 'prove' meaning 'test the limits of,' not 'verify the truth of.'
- Formulate and prove analogous results to provide "evidence" for the validity of the original conjecture.

In addition, every mathematician must acquire various meta-mathematical skills, such as:

- Take a poorly or incompletely posed problem and formulate precise statements to be studied.
- "Fiddle" with a problem, try this-and-that-and-the-other, until eventually some of the ideas that didn't work suddenly fit together to give the solution.

This last item is, in some sense, the most important lesson for a student to absorb. To return to our earlier analogy, if you want to be a writer, the first requirement is that you sit down and write. What you write doesn't have to be good, nor even grammatical, but you have to get some words down on paper. (The modern reader may substitute "computer screen" for piece of paper, although I think it is still true that for mathematical scribbling, nothing beats pencil and paper.) It's not enough to stare at a piece of paper, nor to read other people's novels. Similarly, the first requirement for doing mathematics is to work on it, even if the work doesn't seem to be getting anywhere. You can't *do* mathematics by staring at a blank piece of paper, nor by reading a textbook, nor by listening to a professor explain mathematics on a blackboard.

The books under review appear to have two principal goals. First, they attempt to provide a framework (or a collection of frameworks) on which students can hang the pieces of their proofs. Second, they try to force the students to do mathematical research, by which I mean to solve problems and to write proofs that are not exactly like problems and proofs they've seen before. In short, students who successfully read these books will have to think for themselves, a hard and painful process required for doing mathematics. Beyond that, the three books have different contents and strengths and weaknesses. Which book is to be preferred depends largely on the teacher's tastes and the expected mathematical ability of the students. Herewith are some remarks to help in making that choice.

Exner's book has the heaviest concentration of proof techniques and the lightest load of mathematical content. This is not meant as a criticism, because the techniques it stresses are fundamental, and it explains them in a way designed to grab and hold the students' attention. Virtually every page contains a signal for the student to stop reading and to do some personal work, starting with Exercise 1.1: "Ahem . . . why don't *you* pick a function and we'll see [if it's injective]" and ending with Exercise 3.116: "Take a shot at proving this [the Schröder-Bernstein theorem]." Used in conjunction with a standard text in a traditional first course involving proofs, whether it be algebra, real analysis, or topology, this book will certainly help most students to make the transition from spectator to participant. A lengthy appendix giving exercise hints (but not solutions) will further aid the reader. On the other hand, even with some Laboratories included as Chapter 4, Exner's book would need to be supplemented with additional mathematical content if it were to be used for the full semester transition or bridge course common in many mathematics departments.

Rotman's book contains an interesting blend of miscellaneous mathematical topics interspersed with numerous jokes, stories, and historical anecdotes. The emphasis is on teaching proofs by example using topics that are mathematically interesting but relatively elementary. Thus logic and set theory are largely ignored, since the author "finds this material rather dull and uninspiring, and imagines that this feeling is shared by most students." Some of the more interesting topics include:

- The Pythagorean theorem, Pythagorean triples, and trigonometry.
- Circles, areas, and π.
- Polynomials, including derivation of the formulas for the roots of quadratic, cubic, and quartic polynomials.
- Diophantine approximation and proofs of the irrationality of e and π.

As may be seen from this list, the topics covered are sufficiently elementary that proofs do not require much in the way of definitions or machinery, but they are sufficiently mathematically interesting to appeal to most instructors and most students. However, in contrast to Exner's book, the clearly written proofs do not include much in the way of motivation or discussion of meta-mathematical proof structure, so instructors may need to supply some of this material in class. On the plus side, the proofs are clear and provide excellent models for students who can learn by example, there are lots of exercises of varying degrees of difficulty, and the non-mathematical material is generally fun to read.

The book by D'Angelo and West is the longest and most ambitious of the three books under review. It contains all of the material needed for a transition course aimed at serious mathematics majors. In the first section (74 pages), the authors discuss set theory, logic, and other topics that are used in creating proofs. They then settle down to the main purpose of their book, namely to present several standard areas of mathematics with clear and complete definitions, proofs, and examples. Their topics include number theory, combinatorics, probability, graph theory, and real analysis (of one variable), so everyone will find something of interest. Each chapter ends with a substantial collection of exercises, and an appendix gives hints (but not solutions) for the more difficult exercises.

The three books under review are all worthwhile additions to the textbook market serving somewhat different audiences. Exner's book teaches proof techniques in a bright and lively manner, making it a good supplementary text for a first proof-based course. Rotman's offering provides mathematical content not normally covered in standard mathematics courses; it has carefully written proofs and much entertaining non-mathematical material. It would be especially good for students who hope to study higher mathematics, but who need some coddling before their leap into the cold waters of proof-laden upper-division courses. Finally, the book by D'Angelo and West contains everything needed for a "transition to proofs" course aimed at students with a serious intention of majoring in mathematics.

Brown University, Providence, RI 02912
jhs@math.brown.edu

The Social Life of Numbers: A Quechua Ontology of Numbers and Philosophy of Arithmetic. By Gary Urton (with the collaboration of Primitivo Nina Llanos). University of Texas Press, Austin, 1997, xv + 267 pp., $35.00 hardcover, $17.95 softcover.

Reviewed by **John Meier** and **Trisha Thorme**

1. Introduction. It is probably best to start with a confession: while we are interested observers of ethnomathematics, we are not researchers in the field. John studies geometric group theory, and Trisha is an anthropological archaeologist working in the Andes. However, we co-designed a sophomore-level intensive writing course at Lafayette College called *Counting and Culture* that is essentially a course in ethnomathematics. Although neither one of us is an ethnomathematician, as a pair we come close.

Before developing *Counting and Culture*, we presumed that "ethnomathematics" was the anthropology of mathematical activities, primarily in non-Western cultures. We expected ethnomathematicians to explore the mathematics developed and used by various (non-Western) cultures, the cultural values their mathematical activities carry, and the meanings inherent in them. We were aware of some famous texts in the field, such as *Code of the Quipu*, which basically fit this picture. But as we dug deeper into the literature we found a broad mixture of approaches to mathematics in various cultures. Much of the literature is devoted to educational issues, often promoting the preservation of a particular culture's mathematical heritage. This revitalization of indigenous mathematics is important and worthwhile, and can instill pride in the traditions of that culture. But it is often difficult to see how this is ethnomathematics; rather it seems to be a branch of the history of mathematics that laudably is not focused on European issues, and is interested in pedagogical uses. We also found several articles that catalog mathematical activity in traditional cultures without exploring the cultural context of the mathematics. We were, and still are, disappointed in the small number of articles and books that take a serious anthropological look at mathematics.

Producing catalogs at the expense of context is a general problem in the ethnosciences. Astounded at the sophisticated scientific achievements of so-called primitive cultures, we fail to ask the more interesting question of how engineering, astronomy, agronomy, and so forth were integrated into the culture in question. For example, while figuring out the underlying structure of ancient Mesoamerican calendars is interesting, such a study should integrate the anthropological richness of the region; without contextual connections, one wonders if it is really the *Mesoamerican* calendar being studied. Knowing that some ancient or contemporary culture uses a concept that we put in the domain of science becomes more interesting when we understand what cultural domains (such as kinship, religion, or economy) give meaning to the concept in that culture. Decontextualizing such studies removes individuals, their culture, and their agency from the picture.

One of the main strengths of *The Social Life of Numbers* is the deep and carefully presented anthropological analysis that refuses to reduce Andean mathematical activity to the formalism of Western mathematics.

Gary Urton is an Andean anthropologist who established himself as an exceptional ethnoscientist in his earlier work on ethnoastronomy. His study of astronomy and calendars in a Quechua-speaking community near Cuzco (Peru) illuminated Andean peoples' close attention to celestial and terrestrial geography [**4**]. Quechua was the language of the ruling Inka in their polyglot empire, which stretched from Ecuador to Chile along the spine of the mountain chain, and millions of people in the contemporary Andean nations of Peru, Bolivia, and Ecuador speak Quechua as their first language.

The Andes provide an excellent opportunity for the study of non-Western knowledge systems. A mathematical favorite is the Inkas' sophisticated accounting system, the knotted string records called khipus (or "quipus" in the old-fashioned spelling). This example of Inka cultural logic differs substantially from the roots of our own material conventions for record keeping; see [**2**]. The study of khipus and other Andean technologies helps us to understand these ancient and contemporary cultures and to ask new and difficult questions about our own.

2. What's in the book. *The Social Life of Numbers* opens with a careful and thoughtful critique of the field of ethnomathematics. For someone interested in the field, or for any mathematician interested in connections between mathematics and the humanities, the chapter on Anthropology and the Philosophy of Arithmetic is required reading. Urton unwinds the various threads that have been woven together into "ethnomathematics", showing how the field has been muddled through lack of care in distinguishing educational, philosophical, and anthropological concerns.

> Rather, it is clear that by the use of the label *ethno*mathematics, what is considered to be susceptible to cultural influence are the *conceptions* of numbers and the *philosophy* of mathematics. In other words, to my mind at least, ethnomathematics is actually concerned with *ethnophilosopy*. The elision of "philosophy" in the word ethnomathematics masks the idea, firmly held by most practitioners of ethnomathematics, that while philosophy may be influenced by culture, mathematics is not. [p. 7, emphasis in original]

Much of *The Social Life of Numbers* would fall under the heading of ethnoarithmetic. That is, Urton studies the philosophy of arithmetic (specifically, ontological concerns) indigenous to Quechua-speaking regions of the Andes.

In the subsection Anthropology's Potential Contribution to Mathematics, Urton directly addresses the troubling question of how ethnomathematics can contribute to mathematics. Clearly ethnomathematics can contribute to anthropology's goal of studying the diversity of human culture, but Urton also pursues the converse: Mathematicians ought to take more than a casual look at the anthropology of their field.

If ethnomathematics were to contribute to mathematics, one would generally expect the "contribution" to be some new relation previously unknown in the West. Urton suggests that there are other ways ethnomathematics can be of service. If mathematical philosophies vary across cultures, can an anthropological study of how these philosophies integrate mathematics into a cultural context help mathematicians clarify the philosophical grounding of their field? It is a very good question.

While a phrase such as "mathematics is a cultural construct" may not raise an eyebrow amongst our colleagues in the humanities, it does directly contravene mathematicians' inherent Platonism. As Urton states (p. 17), "the view that mathematical truths do not—*cannot*—differ cross-culturally is a central tenet of

the Platonist vision of the reality of numbers, shapes, sets, etc." While the neo-Platonic viewpoint of most mathematicians has come under attack in the philosophy of mathematics, perhaps an anthropological approach will have a greater impact on how we do mathematics. Our attachment to Platonism is almost surely more cultural than philosophical.

Urton explores the meaning of numbers and arithmetical operations in Quechua-speaking communities. Instead of taking as given the universality of numbers and counting, he looks at the origin and culturally specific significance of numbers. This ontology reveals the social nature of numbers, counting, and sets. Through examining ordinal sequences such as ears of corn, the tines of a pitchfork, and most importantly, a mother and her offspring, Urton concludes (p. 95) that "it is only by coupling the number series with the organizing forces of social relations that ordinal sequences emerge as meaningful, rational expressions of the relations and forces organizing and uniting the members of a group." This discovery of relatedness as the driving force of numbers and sets is one of the most important points in the text.

This leads to Urton's explication of the creation and maintenance of balance and harmony as the underlying rationale for arithmetical operations. He terms this *rectification.*

> Simply stated, rectification is grounded in the idealistic, philosophical, and cosmological notion that all things in the world—from material objects to habits, attitudes, and emotions—ought to be in a state of balance, or equilibrium.... In the contexts that we are interested in, the major forms of rectification include addition, subtraction, multiplication, and division. While each of these concepts and practices has a variety of labels in different mathematical contexts, collectively they are grouped together as corrective actions aimed—each in a slightly different way from the others—at achieving rectification in a circumstance of imbalance, disequilibrium, and disharmony. [pp. 145–146]

More examples, such as weaving and the cultural meaning of odd and even numbers, further illustrate this idea. This insight accords well with what we know about Andean society, particularly the paradigm that balance is created out of the melding of opposites.

In the penultimate chapter, Urton turns his eye toward historical and archaeological sources. He makes a convincing case that one can see the "arithmetic of rectification" and the association between numbers and the model of a mother with her age-graded offspring (or at the very least, the central role of sets of five) in ancient Andean civilizations. He rightly views the conquest of the Inka by the Spanish as an important test case for fairly radical ideas of Bishop who argues that mathematics was an important tool in the imposition of Western culture [**3**]. Urton presents convincing evidence that this was indeed the case for the Spanish conquest.

> ...(M)uch of what was initially imposed and subsequently became embedded as elements of everyday practices of colonial administration in the Americas rode in on the back of a new political arithmetic that gave coherence and authority to the Spanish colonial adventure. [p. 196]

It is clear from studying khipus that the Inka used a base ten number system. So it is not the number system itself that was imposed by the Spanish, rather it was the symbols used by the Spanish, and more importantly, the Spanish philosophy of arithmetic.

A striking piece of evidence of this imperialism can be found in a collection of drawings of Inka kings dressed in tunics decorated with Hindu-Arabic numerals.

This clearly indicates not only Spanish influence, since such designs did not exist before the invasion, but also the importance of these symbols. "After, perhaps, the God of Christianity, numbers—that is, the symbols that were used by Spanish administrators for keeping census and tribute records—represented one of the most powerful instruments of colonial rule in the Andes" (p. 205). Urton concludes by arguing that one of the most important, and yet quite subtle, conflicts that occurred was between Andean and Spanish conceptions of arithmetic operations.

3. Critique. *The Social Life of Numbers* is an important contribution to anthropology, mathematics, and ethnomathematics. By cracking open the very definition of ethnomathematics, Urton has clarified muddled issues and directly addressed the vexing issue of why either of the fields from which ethnomathematics is mixed should care about results of the mixture. This case study in the anthropology of number philosophy and ethnoarithmetic does not lose sight of the anthropological context of the mathematics. It is also an important addition to the ethnography of the Andes.

Regrettably, we can't claim this is a flawless text. As Urton confesses, *The Social Life of Numbers* is weak on the mathematics. His arguments are primarily concerned with arithmetic, certainly an important and often neglected aspect of mathematics, but certainly not the only mathematics developed in the Andes. Other accessible examples can be found in the geometry of textiles and architecture. If one is to look for ontological foundations for arithmetic within a culture, it is reasonable to believe that such foundations would be equally interesting for other aspects of mathematics. Urton anticipates this criticism:

> ...(A)ny mathematically sophisticated reader who has stayed with me to this point must be asking: "So, where's the mathematics?" Just so that that reader will not be under the impression that I think that there *is* anything mathematically complex or fancy about the foregoing, I want to state clearly that I do not think that. I leave it to those who are truly mathematically inclined, *and who have a firm grasp of the Quechua language and culture as experienced in the Andes*, to expand on the information and issues raised in this study... in order to illuminate and advance our knowledge about what I firmly believe to be a true depth and complexity in Quechua numerical ideas and arithmetic and mathematical practice. [p. 214, emphasis in original]

A related criticism is that this book is only the first half of a larger project. "I must admit that I did not take up this study from a profound, single-minded interest in numbers and mathematics" (p. 1). Urton is ultimately interested in a study of khipus as proto-writing; this book was in some sense forced upon him because the field of ethnomathematics had not prepared the ground for the latter study. We suspect that Urton's insights into khipus will be of more interest to a general reader and will provide further evidence for the importance of understanding Andean philosophies of arithmetic.

Mathematicians will be happy that Urton's writing is never overburdened with jargon, and that he provides good general insights and commentaries throughout the text. For example, after describing the use of the fingers as a counting tool in the Quechua (for example) base ten numerical system, he cautions: "But we should not allow the fact of the ubiquity of the relationship between fingers and numbers to lull us into seeing this as a trivial relationship, for we will find in these data explicit links among numbers, the classification and organization of the fingers, and ideas about human reproduction, age, gender, and kinship relations" (p. 73). Mathematicians will also be impressed by how well Urton has absorbed the history and philosophy of mathematics.

> To the degree that the introduction into European mathematics of the Hindu-Arabic numeral for zero precipitated a philosophical and theological crisis, at the same time that it allowed (if not sparked) tremendous advances in abstract mathematics, it appears that the Quechua do not today, nor did they in the past, present themselves with the conditions for such a crisis by *naming* the empty set with a cardinal number. We are no doubt also confronted here with one of the implications of the absence of a system of written numerals in the Andes—that is, of *making* a mark to indicate nothingness in a system of writing signs on a two-dimensional surface, as opposed to *not making* any knots in a piece of string to indicate the same (absence of) value. [p. 50]

On the other hand, the book's organization is weak. For mathematicians used to the orderly development of ideas, this can be quite annoying. For example, after discussing the complicated number memory necessary for being a *Mama*—a master weaver—Urton relates how weavers receive their pay and immediately hand it to someone else to count. While the incredible thread manipulations involved in weaving prove that Mamas can count, they cannot count *money* (p. 136). This observation relates to the theme of what is countable and what is not, a main part of the book to this point, but also to the historical subjugation of the Andean region after the Spanish invasion. This historical dimension, which involved the introduction of money, and thus the imposition of "a new set of values that lay behind the conception of numbers and the manipulation of those numbers in arithmetic and mathematical operations" (p. 196), is addressed much later.

Of course, anytime one is sufficiently bold as to write for two different audiences, there are some criticisms on the proportions used in the blending. On the one hand, Urton's mode of argument may be a problem for a mathematician. Since the field does not lend itself to airtight proofs, Urton relies on the standard anthropological technique of illustrating his concepts in various contexts. For example, the numerous illustrations of "mother as the origin of numbers" may tire mathematicians but are sure to satisfy anthropologists. On the other hand, anthropologists are not used to giving arithmetic this much thought.

In total, this book is an important contribution to ethnomathematics, both in the research presented and in Urton's careful analysis of the field. We look forward to his study of khipus and further insights about the indigenous mathematics of the Andes.

REFERENCES

1. Marcia Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks/Cole, 1991.
2. Marcia Ascher and Robert Ascher, *Code of the Quipu: A Study in Media, Mathematics and Culture*, University of Michigan Press, 1981; reprinted as *Mathematics of the Incas: Code of the Quipu*, Dover, 1997.
3. A. Bishop, Western Mathematics: The Secret Weapon of Cultural Imperialism, *Race and Class* **32** (1990) 51–65.
4. Gary Urton, *At the Crossroads of the Earth and the Sky: An Andean Cosmology*, University of Texas Press, 1981.

Lafayette College, Easton, PA 18042
meierj@lafayette.edu

Department of Anthropology, Cornell University, Ithaca, NY 14853
tat2@cornell.edu

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee**
with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	*P : Professional Reading*	*1–4: Semester*
C : Computer Software	*L : Undergraduate Library*	*** : Special Emphasis*
S : Supplementary Reading	*13: Grade Level*	*?? : Questionable*

Readers are advised that price information is subject to change. Selected books receive a second, more extensive review in the *Monthly*.

Books submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

General, S*(13–16), L.** *The Language of Mathematics: Making the Invisible Visible.* Keith Devlin. WH Freeman, 1998, viii + 344 pp, $24.95. [ISBN 0-7167-3379-X] A "spin-off" (with far fewer glossy illustrations) of Devlin's masterful *Mathematics: The Science of Patterns* (TR, October 1997) aimed at a "much wider" audience. This worthy successor to Courant and Robbins' 1941 classic *What is Mathematics?* adds two new chapters on patterns of chance and of cosmology. Strangely, there are absolutely no references or suggestions for related reading. LAS

Mathematics Appreciation, S(15–16), L. *The Jungles of Randomness: A Mathematical Safari.* Ivars Peterson. Wiley, 1998, xiii + 239 pp, $24.95. [ISBN 0-471-16449-6] The latest in Peterson's guided tours of mathematics. Beautifully produced, and very reader-friendly. BC

Logic, T(15–16: 1), L. *Logic for Mathematics and Computer Science.* Stanley N. Burris. Prentice Hall, 1998, xix + 420 pp. [ISBN 0-13-285974-2] Covers traditional logic (propositional logic; syntax, semantics and completeness of first-order logic), equational logic, and computational logic (resolution, unification, term rewrite systems, and skolemization). Includes historical remarks. Does not assume background in analysis or abstract algebra. KES

Foundations, T(16–17: 1), S, P, L. *A Course on Borel Sets.* S.M. Srivastava. Grad. Texts in Math., V. 180. Springer-Verlag, 1998, xvi + 261 pp, $49.95. [ISBN 0-387-98412-7] Introduction to Borel sets and measurable selections. First two chapters provide all necessary set-theoretic and topological background. Avoids use of forcing, large cardinals, games, constructibility, and recursive function theory. Exercises integral to text; often assumed later as theorems. LB

Foundations, T(13–14: 1), L. *An Introduction to Mathematical Reasoning: Numbers, Sets and Functions.* Peter J. Eccles. Cambridge Univ Pr, 1997, xii + 350 pp, $24.95 (P); $59.95. [ISBN 0-521-59718-8; 0-521-59269-0] Well-written text for transition course. Covers basic symbolic logic, methods of proof, finite and infinite sets, counting principles, elementary number theory. Proofs of many theorems preceded by paragraph on constructing the proof. Complete solutions for exercises. KES

Foundations, S(16–17). *A Primer of Abstract Mathematics.* Robert B. Ash. MAA, 1998, x + 181 pp, $27.95 (P). [ISBN 0-88385-708-1] Very concise presentation of concepts and results in counting, number theory, set theory, linear algebra, linear operators. Exercises with complete solutions. Suggested audience: graduate students or professionals who need quick review of (or introduction to) these topics. KES

Combinatorics, T*(13–16: 1), S, L. *Combinatorics: A Problem Oriented Approach.* Daniel A. Marcus. MAA, 1998, x + 136 pp, $28 (P). [ISBN 0-88385-710-3] An easy route into combinatorics through problem solving. Use first 50–75% in first/second year discrete math course, or with some supplementation in junior/senior combinatorics. Gets as far as generating functions and Polya–Redfield theory. DB

Discrete Mathematics, T(17–18: 1, 2), L**.**

Modern Graph Theory. Béla Bollobás. Grad. Texts in Math., V. 184. Springer-Verlag, 1998, xiii + 394 pp, $59.95 (P). [ISBN 0-387-98491-7] More than twice the length of *Graph Theory: An Introductory Course*, from which it evolved, this up-to-date and in-depth introduction contains much recent material, including Szemerédi's Regularity Lemma, list colorings, phase transition in random graphs, random walks on graphs, the Tutte polynomial and its relatives in knot theory. Many exercises, graded by difficulty. Clearly written. No extensive prerequisites, but does require graduate-level mathematical maturity. LB

Discrete Mathematics, S(17–18: 1, 2), L. *Exercises in Graph Theory.* O. Melnikov, *et al.* Texts in Math. Sci., V. 19. Kluwer Academic, 1998, viii + 354 pp, $149. [ISBN 0-7923-4906-7] Over a thousand exercises to accompany the authors' *Lectures on Graph Theory*, with answers, hints, and solutions. Organized in same order and uses same notation as text. Trees, independence and coverings, matchings, tours, planarity, colorings, degree sequences, connectivity, digraphs, hypergraphs. Could supplement many other introductory texts. LB

Number Theory, T(16–18: 1), S, P, L. *Elliptic Curves: Function Theory, Geometry, Arithmetic.* Henry McKean, Victor Moll. Cambridge Univ Pr, 1997, xiii + 280 pp, $59.95. [ISBN 0-521-58228-8] An attractive, well-motivated presentation of the theory of elliptic curves, integrals, and functions. BC

Number Theory, T(16–17: 1), S, P, L. *Algorithmic Algebraic Number Theory.* M. Pohst, H. Zassenhaus. Ency. of Math. & Its Applic., V. 30. Cambridge Univ Pr, 1997, xiv + 499 pp, $39.95 (P); $115. [ISBN 0-521-59669-6; 0-521-33060-2] How do you actually compute a Galois group? Or the group of units, or the class group, of an algebraic number field? Answers these and other computational (and theoretical) questions. (1989 text edition, TR, August–September 1990.) BC

Number Theory, S(17–18), P. *Automorphic Forms on $SL_2(\mathbb{R})$.* Armand Borel. Tracts in Math., V. 130. Cambridge Univ Pr, 1997, x + 192 pp, $47.95. [ISBN 0-521-58049-8] An introduction to the analytic theory of automorphic (and cusp) forms, with chapters on Eisenstein series and spectral decomposition. BC

Linear Algebra, T(14: 1), C. *Interactive Linear Algebra with MAPLE V.* Elias Deeba, Ananda Gunawardena. Springer-Verlag, 1998, xiii + 317 pp, $49.95 (P), with CD–ROM. [ISBN 0-387-98240-X] Interactive computer lessons with exercises, laboratories, and an adaptive testing system, supported by a GUI. Text is an edited version of the interactive lessons (which are on the CD–ROM). PG

Linear Algebra, T(14: 1). *Linear Algebra with Applications.* George Nakos, David Joyner. Brooks/Cole, 1998, xviii + 666 pp, $54. [ISBN 0-534-95526-6] Over 400 examples. Emphasizes geometric insight, numerical methods, and a renewed interest in determinants. Each chapter ends with "miniprojects" and computer exercises. Standard ordering of topics. TH

Linear Algebra, T(13–14). *Introduction to Linear Algebra, Second Edition.* Gilbert Strang. Wellesley–Cambridge Pr, 1998, viii + 503 pp, $62.50. [ISBN 0-9614088-5-5] The *Second Edition* for this excellent linear algebra text with theory and applications. (*First Edition*, TR, January 1994.) PF

Ring Theory, P. *Algebras and Modules, I & II.* Eds: Idun Reiten, Sverre O. Smalø, Øyvind Solberg. AMS, 1998. *I*, Canadian Math. Soc. Conf. Proc., V. 23, xiv + 198 pp, $39 (P), [ISBN 0-8218-0850-8]; *II*, Canadian Math. Soc. Conf. Proc., V. 24, xxi + 569 pp, $99 (P). [ISBN 0-8218-1076-6] Proceedings of 1996 workshops in Trondheim and Geiranger, Norway.

Algebra, T(17–18: 1), S, L. *Categories for the Working Mathematician, Second Edition.* Saunders Mac Lane. Grad. Texts in Math., V. 5. Springer-Verlag, 1998, xii + 314 pp, $54.95. [ISBN 0-387-98403-8] Adds two chapters: one on monoidal categories and coherence theorems; one on 2–categories and higher-dimensional categories. Updated bibliography; minor revisions to earlier chapters. (*First Edition*, TR, June–July 1972.) LB

Algebra, T(18: 1), P. *Linear Algebraic Groups, Second Edition.* T.A. Springer. Progress in Math., V. 9. Birkhäuser Boston, 1998, xi + 334 pp, $64.95. [ISBN 0-8176-4021-5] Thoroughly revised; extends theory to arbitrary fields. (*First Edition*, TR, April 1982.) TH

Algebra, T(18: 1, 2), P. *Character Theory of Finite Groups.* Bertram Huppert. Expos. in Math., V. 25. Walter de Gruyter, 1998, vi + 618 pp, $328. [ISBN 3-11-015421-8] Self-contained treatment with recent achievements (especially in degree problems, conjugacy class length). Numerous examples. TH

Algebra, T(18: 1, 2), P. *Characters of Finite Groups. Part 2.* Ya. G. Berkovich, E.M. Zhmud'. Transl. of Math. Mono., V. 181. AMS, 1999, xxv + 332 pp, $115. [ISBN 0-8218-0532-0] Second of a two-part text on characters of

representations of finite groups over complex fields. TH

Algebra, P. *Algebra in the Stone–Čech Compactification: Theory and Applications.* Neil Hindman, Dona Strauss. Expos. in Math., V. 27. Walter de Gruyter, 1998, xiii + 485 pp, $128.95. [ISBN 3-11-015420-X] The algebra of the Stone–Čech compactification of a discrete semigroup and its applications to Ramsey theory and topological dynamics. DB

Calculus, T*(13–14), S, L.** *Basic Calculus: From Archimedes to Newton to its Role in Science.* Alexander J. Hahn. Springer-Verlag, 1998, xiv + 545 pp, $59.95. [ISBN 0-387-94606-3] Rich in examples of how calculus has been/is being used. Would be difficult as a text: loses forest for the trees. Suitable for honors course with students who had calculus in high school, know the techniques, seek deeper understanding. DB

Calculus, S(13). *Calculus.* Elliot C. Gootman. Coll. Rev. Ser. Math. Barron's Educational Ser, 1997, viii + 342 pp, $13.95 (P). [ISBN 0-8120-9819-6] Supplement for a standard calculus course. Algebra and geometry reviews distributed throughout the text. PF

Differential Equations, T(17), P, L. *Ordinary Differential Equations.* Wolfgang Walter. Transl: Russell Thompson. Grad. Texts in Math., V. 182. Springer-Verlag, 1998, xi + 380 pp, $49.95. [ISBN 0-387-98459-3] Rigorous development of the theory of ordinary DEs. Uses holomorphic functions, functional analysis, and differential inequalities. Supplemental sections use Lebesgue theory and Schauder's fixed point theorem. PG

Differential Equations, T(14: 1), L. *Differential Equations with Computer Lab Experiments, Second Edition.* Dennis G. Zill. Brooks/Cole, 1998, xiii + 463 pp, $80.95. [ISBN 0-534-35173-5] Begins with qualitative study of DEs. Topics: first- and higher-order DEs, systems, phase portraits and stability, Laplace transforms, and series solutions. Routine exercises, computer labs, projects, discussion problems, and applications. (*First Edition,* TR, November 1995.) PG

Partial Differential Equations, T(16: 1), L. *Introduction to Partial Differential Equations: A Computational Approach.* Aslak Tveito, Ragnar Winther. Texts in Appl. Math., V. 29. Springer-Verlag, 1998, xv + 392 pp, $54.95. [ISBN 0-387-98327-9] Thorough treatment of Poisson's equation in one- and two-dimensions (including discussions of a Green's function, finite difference approximation, and general eigenvalue problems). Treats one-dimensional heat and wave equations similarly. Discusses convergence of Fourier series. Includes exercises and projects. PG

Partial Differential Equations, T(17: 1), C, P, L. *Partial Differential Equations for Computational Science: With Maple and Vector Analysis.* David Betounes. Springer-Verlag, 1998, xix + 517 pp, $64.95, with CD–ROM. [ISBN 0-387-98300-7] Sophisticated introduction to PDEs. Emphasizes use of computer for visualization and interpretation of solutions to heat, wave, and Poisson's equations. Also discusses Sturm–Liouville problem and systems (e.g., Maxwell's equations and Navier–Stokes equations). PG

Numerical Analysis, P, L. *Methods for Solving Systems of Nonlinear Equations, Second Edition.* Werner C. Rheinboldt. CBMS–NSF Reg. Conf. Ser. in Appl. Math., V. 70. SIAM, 1998, ix + 148 pp, $34 (P). [ISBN 0-89871-415-X] Extensively revised and expanded (*First Edition,* TR, April 1975). Additions include an introduction to the theory of inexact Newton methods, basic theory of continuation methods, and new material on minimization methods. AO

Numerical Analysis, T(15: 1), C, L. *Numerical Methods, Second Edition.* J. Douglas Faires, Richard Burden. Brooks/Cole, 1998, xi + 594 pp, $60, with disk. [ISBN 0-534-35187-5] Emphasizes the application of approximation techniques to problems from engineering, physical, and computer sciences. Motivates and describes techniques from an implementation standpoint. Uses Maple. PG

Numerical Analysis, P. *Domain Decomposition Methods 10.* Eds: Jan Mandel, Charbel Farhat, Xiao-Chuan Cai. Contemp. Math., V. 218. AMS, 1998, xii + 554 pp, $110 (P). [ISBN 0-8218-0988-1] Proceedings of a 1997 conference held at the University of Colorado at Boulder.

Analysis, T(18), P. *Wavelets: Theory and Applications.* A.K. Louis, P. Maaß, A. Rieder. Wiley, 1997, xvii + 324 pp, $99.75. [ISBN 0-471-96792-0] Translation of 1994 German edition. Continuous and discrete wavelet transforms, applications to signals, data compression, boundary value problems. SN

Analysis, T(16–17: 1, 2), L. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.* Robert Goldblatt. Grad. Texts in Math., V. 188. Springer-Verlag, 1998, xiv + 289 pp, $49.95. [ISBN 0-387-98464-X] Accessible introduction to nonstandard analysis. Many examples. Good for those without formal logic background. Rigorous construction

of the hyperreals followed by development of one-variable calculus, analysis, and topology from the nonstandard perspective. SN

Geometry, T(15–17). *Projective Geometry: From Foundations to Applications.* Albrecht Beutelspacher, Ute Rosenbaum. Cambridge Univ Pr, 1998, x + 258 pp, $24.95 (P); $64.95. [ISBN 0-521-48364-6; 0-521-48277-1] An axiomatic approach to projective geometry with many applications throughout. PF

Geometry, T(15–17), L. *Foundations of Convex Geometry.* W.A. Coppel. Australian Math. Soc. Lect. Ser., V. 12. Cambridge Univ Pr, 1998, xiv + 222 pp, $39.95 (P). [ISBN 0-521-63970-0] A coordinate-free, axiomatically evolving treatise on geometry. PF

Operations Research, P, L. *Operations Research in Transportation Systems: Ideas and Schemes of Optimization Methods for Strategic Planning and Operations Management.* Alexander S. Belenky. Appl. Optim., V. 20. Kluwer Academic, 1998, viii + 440 pp, $195. [ISBN 0-7923-5157-6] Survey of optimization methods including linear and nonlinear programming, integer programming, network flow problems, stochastic programming, optimal control, and game theory. Each chapter concludes with an extensive bibliography. AO

Optimization, T(15–16), S, L. *Cake-Cutting Algorithms: Be Fair If You Can.* Jack Robertson, William Webb. AK Peters, 1998, x + 181 pp, $38. [ISBN 1-56881-076-8] An elementary introduction to algorithms for fair division of any homogeneous divisible "cake." Six informal chapters introduce problems and strategies; a seventh summarizes results; the final four provide proofs. Exercises (with solutions at the back) and projects make this nice monograph ideal for a seminar. LAS

Optimization, P. *Convex Analysis and Global Optimization.* Hoang Tuy. Nonconvex Optim. & Its Applic., V. 22. Kluwer Academic, 1998, xi + 339 pp, $136. [ISBN 0-7923-4818-4]

Probability, T(16–17), S, P, L. *Integral, Probability, and Fractal Measures.* Gerald A. Edgar. Springer-Verlag, 1998, x + 286 pp, $39.95. [ISBN 0-387-98205-1] A continuation of the author's *Measure, Topology, and Fractal Geometry.* Gives a solid, measure-theoretic grounding for the study of fractals, including random fractals. BC

Statistics, P. *Advances in Stochastic Models for Reliability, Quality and Safety.* Eds: Waltraud Kahle, *et al.* Stat. for Industry & Tech. Birkhäuser Boston, 1998, xxvii + 382 pp, $79.95. [ISBN 0-8176-4049-5] 24 papers from a 1997 workshop held in Magdeburg, Germany. In 4 parts: Lifetime Analysis; Reliability Analysis; Network Analysis; Process Control.

Computer Science, T?(14–16: 1), S, P. *Purely Functional Data Structures.* Chris Okasaki. Cambridge Univ Pr, 1998, x + 220 pp, $39.95. [ISBN 0-521-63124-6] Conceptual discussion, and interesting techniques, for implementing data structures in functional languages. Good discussion of distinctions between imperative (ephemeral) and functional (persistent) structural features, role of lazy evaluation, amortization, lazy rebuilding, data structure bootstrapping. Examples in Standard ML, with Haskell appendix. RM

Applications (Biological Science), S(16–17), P, L. *An Introduction to Structured Population Dynamics.* J.M. Cushing. CBMS–NSF Reg. Conf. Ser. in Appl. Math., V. 71. SIAM, 1998, xiii + 193 pp, $36 (P). [ISBN 0-89871-417-6] Theory and applications of discrete and continuous structured population models. (Models which allow some characteristics of individual organisms to be incorporated into the study of population dynamics.) Focuses on asymptotic dynamics of deterministic models. AO

Applications (Engineering), T(15: 2), L. *Advanced Engineering Mathematics.* Dean G. Duffy. CRC Pr, 1998, 634 pp. [ISBN 0-8493-7854-0] Begins with discussion of complex variables, providing background for Fourier, Laplace, and z-transforms. Includes topics in multivariable calculus (e.g., Sturm–Liouville problem, heat, wave, and Laplace equations). Concludes with vector calculus (e.g., surface integrals, Stoke's theorem, and the use of eigenvalues in ODEs). PG

Applications, T?(15–16), S*, P, L. *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation.* Gary William Flake. MIT Pr, 1998, xviii + 493 pp, $45. [ISBN 0-262-06200-3] Intriguing, in-depth overview of the role of computer explorations and simulation for understanding complex natural phenomena. Emphasis on fractals and chaos, with the theme that the computational properties of interactions underlie notions of beauty and interestingness of nature. Web resources available for simulation software, etc. RM

Reviewers

LB: Lynne Baur, Carleton; DB: David Bressoud, Macalester; BC: Barry Cipra, St. Olaf; PF: Paul Froeschl, Macalester; PG: Philip Gloor, St. Olaf; TH: Tom Halverson, Macalester; RM: Richard Molnar, Macalester; SN: Sam Northshield, Carleton; AO: Arnold Ostebee, St. Olaf; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf.

EDITOR'S ENDNOTES

Alfred Manaster wrote about his article, Some Characteristics of Eighth Grade Mathematics Classes in the TIMSS Videotape Study, this MONTHLY 105 (1998) 793-805:

> Professor Henry Adler was kind enough to point out an error in each use of the symbol χ^2 to indicate a value in my article. In each case, the value refers to the probability, usually denoted by p, that the value of the chi-square statistic would be at least as large as the observed value under the hypothesis that there is no relationship between the response and the factor. I am responsible for this error and hope that it has not confused any readers.

Gerald A. Heuer shared with us an email message to Marc Frantz about the latter's article, Two Functions Whose Powers Make Fractals, this MONTHLY 105 (1998) 618-630:

> In your article in the August MONTHLY you refer to the differentiability of the function (there called f) sometimes called the "ruler" function, and make the statement: "Darst and Taylor showed that if $1 \leq p \leq 2$, then f^p is nowhere differentiable, and if $p > 2$, then f^p is differentiable almost everywhere." Darst and Taylor did indeed show this in their 1996 MONTHLY article, but the result is much older. G. J. Porter, this MONTHLY 69 (1962) 142, showed that f itself is nowhere differentiable, and in the article Functions Continuous at the Irrationals and Discontinuous at the Rationals, this MONTHLY 72 (1965) 370-373, some undergraduates and I proved that if $0 < p < 2$, then f^p is nowhere Lipschitzian (implying nowhere differentiable), that f^2 is nowhere differentiable, that for $p > 2$, f^p is almost everywhere differentiable, and somewhat more. The results were extended further in the note, A Property of Functions Continuous on a Dense Set, this MONTHLY 73 (1966) 378-379.
>
> At the time, I got some correspondence from Solomon Marcus indicating that he had proved some related results (I believe overlapping ours) earlier....Of course, as we both know, people are rediscovering old results all the time, often without knowing of their earlier establishment.

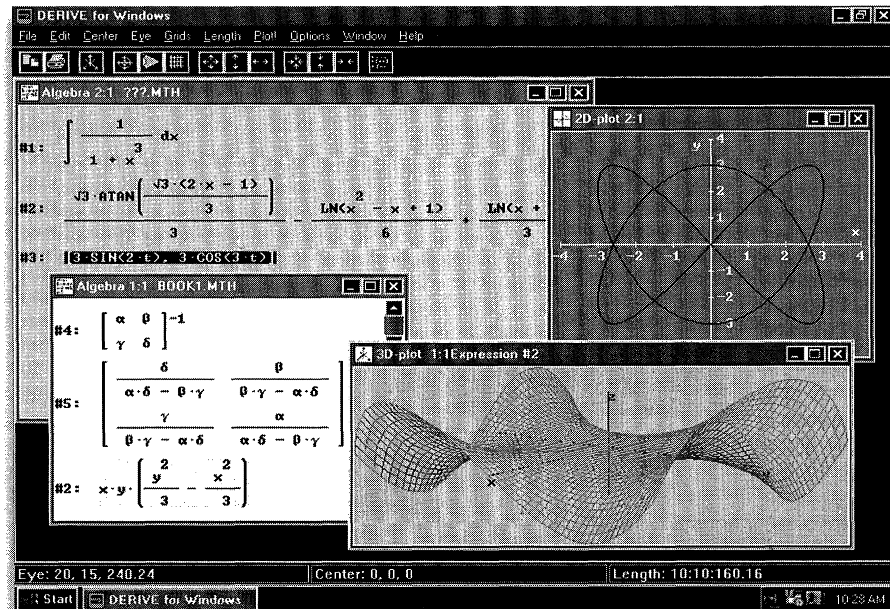
In his email response, Frantz wrote:

> It looks as if the subject has an interesting history. I find it particularly ironic that all the references you mentioned were from the same journal! It makes me think that someday there should be a computerized system, more thorough and easy to use than anything available now, which would help authors (and referees) avoid duplication of results and give more thorough references. Until then we'll have to settle for being educated after the fact.

Well, Marc (and all our authors, referees, and readers) that day is almost here. If all goes well, sometime in 1999, all of the MONTHLY from Vol. 1 in 1894 up to five years ago (this cutoff rolls forward each year) will be available online at JSTOR as graphic images and also in fully-searchable form. Watch MAA Online for an announcement and details when this exciting new service becomes available. Meanwhile, you can learn more about the Andrew W. Mellon Foundation's Journal Storage Project at www.mellon.org. If you visit www.jstor.org and click on "Demo", you can experiment with searching and viewing a demonstration database of three journals; clicking on "About the Full JSTOR Collection" gives access to detailed information about all aspects of JSTOR, including a list of participating journals...it is a pleasure to note that the MONTHLY is now among the latter.

Roger A. Horn, *Editor*

NEW! ***Site Licenses and Student Pricing.***
See www.derive.com



DERIVE® for Windows

DERIVE is the trusted mathematical assistant relied upon by students, educators, engineers, and scientists around the world. It does for algebra, equations, trigonometry, vectors, matrices, and calculus what the scientific calculator does for numbers — it eliminates the drudgery of performing long and tedious mathematical calculations. You can easily solve both symbolic and numeric problems and see the results plotted as 2D or 3D graphs.

For everyday mathematical work *DERIVE* is a tireless, powerful, and knowledgeable assistant. For teaching or learning mathematics, *DERIVE* gives you the freedom to explore different mathematical approaches better and more quickly than by using traditional methods.

System Requirements: Windows 95, 3.1x or NT running on a computer with 8 megabytes of memory.

Suggested Retail Price: $250. Educational pricing available.

For product information and list of dealers, fax, email, write, or call Soft Warehouse, Inc. or visit our website at http://www.derive.com.

The Easiest just got Easier.

Soft Warehouse Inc.
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waialae Avenue
Suite 304 • Honolulu, Hawaii, USA 96816-3259
Telephone: (808) 734-5801 after 10:00 a.m. PST
Fax: (808) 735-1105 • Email: swh@aloha.com.

THE MATHEMATICAL ASSOCIATION OF AMERICA

Magic Tricks, Card Shuffling, and Dynamic Computer Memories

S. Brent Morris

Series: Spectrum

Magic Tricks, Card Shuffling, and Dynamic Computer Memories is a book that explores the fascinating interconnections between these seemingly unrelated topics. It is written for undergraduate mathematics, computer science, and electrical engineering majors, but it is accessible to motivated high school math students and magicians who want to understand the mathematics of card shuffling. It is a fun book that stands alone, but it could nicely supplement classes in discrete mathematics, combinatorics, algorithms, and computer networks. This book looks at the mathematics of the perfect shuffle and develops the algorithms for controlling dynamic memories (and doing some clever card tricks).

Each chapter begins with the description of a card trick and ends with its explanation, usually using mathematics developed in the chapter. The book itself is designed as a prop for a trick, but you don't need to use or understand any of its mathematics to do some good magic.

Read what reviewers have said about this book!

***Magic Tricks, Card Shuffling, and Dynamic Computer Memories** is essential reading for any magic buff who can faro shuffle or who wishes to acquire this unusual skill. The book will also be of great interest to computer scientists and to mathematicians working in the field of combinatorics ... anyone can read it with enjoyment and profit who is curious about the art and mathematics of card magic, or about the unexpected application of perfect shuffles to the storage and retrieval of computer information.*

—Martin Gardner

Provides a fascinating mix of history, mathematics and great magic tricks. I learned something on every page.

—Ron Graham, Chief Scientist, AT&T

....a tour de force for Morris... This is a most unusual and effective way to learn the concepts embodied in the interconnections of today's parallel computers.

—Harold Stone, NEC Research, Princeton

Table of Contents: **The Perfect Shuffle:** The Origins of the Perfect Shuffle; The Faro Dealer's Shuffle; The Mathematical Model of the Perfect Shuffle; The Stay-Stack Principle; Trick 1.2 (The Seekers, by Paul Swinford); **The Order of Shuffles:** The Product of Shuffles; Moving a Card in a Deck; Trick 2.9 (A Spelling Bee); **Shuffle Groups:** Randomizing a Deck of Cards; Shuffles and Cuts in Even Decks; Shuffles and Cuts in Odd Decks; Out- and In-Shuffles in an Even Deck; Trick 3.8 (A Challenge Poker Deal); **Generalizing the Perfect Shuffle:** Out-Shuffling Several Packets of Cards; Looking for a Neat Formula; Permutation Matrices; Generalizing Theorems; Generalized Shuffle Groups; Generalizing the In-Shuffle; Trick 4.7 (The Triple Seekers); **Dynamic Computer Memories:** The Shift-Register Memory; The Perfect Shuffle Memory; The Shift-Shuffle Memory; Details, Details, Details; The Perfect-Shuffle Memory for $N=2n$: Sequential Accessing in a Perfect-Shuffle Memory of Size $N=2n$; Properties of Tours; Epilogue; Trick 5.20. (Unshuffled by Paul Gertner); **Appendix 1:** The Order of Shuffles; **Appendix 2:** How to do the Faro Shuffle; The Double or Ordinary Faro Shuffle; The Triple Faro Shuffle; Appendix 3: Tours on Decks of Size 8, 16, 32, and 64; **Bibliography:** Selected Perfect Shuffle References.

Catalog Code: CARDS/JR
150 pp., Paperbound, 1998
ISBN 0-88385-527-5
List: $28.95 MAA Member: $22.50

Phone in Your Order Now! ☎ *1-800-331-1622*

Monday – Friday 8:30 am – 5:00 pm FAX (301) 206-9789

or mail to: The Mathematical Association of America, PO Box 91112, Washington, DC 20090-1112

Shipping and Handling: Postage and handling are charged as follows: **USA orders (shipped via UPS):** $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. Credit Card orders are accepted for all customers.

Name ______________________

Address ______________________

City ____________ State _____ Zip ________

Phone ______________________

QTY.	CATALOG CODE	PRICE	AMOUNT
______	**CARDS/JR**	______	______

All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.

Shipping & handling ______

TOTAL ______

Payment ☐ Check ☐ VISA ☐ MasterCard

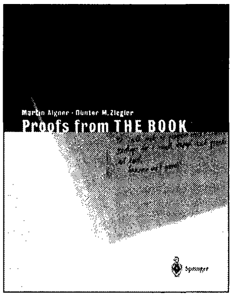
Credit Card No. ______________ Expires ___/___

Signature ______________________

SPRINGER FOR MATHEMATICS

MARTIN AIGNER, Freie Universität Berlin and **GÜNTER M. ZIEGLER,** Technische Universität Berlin, both in Germany

PROOFS FROM THE BOOK



The (mathematical) heroes of this book are "perfect proofs": brilliant ideas, clever connections, and wonderful observations that bring new insight and surprising perspectives on basic and challenging problems from number theory, geometry, analysis, combinatorics, and graph theory. Thirty beautiful examples are presented here. They are candidates for "The Book" in which God records the perfect proofs—according to the late Paul Erdös, who himself suggested many of the topics in this collection. The result is a book which will be fun for everybody with an interest in mathematics, requiring only a very modest (undergraduate) mathematical background.

1998/210 PP., 220 ILLUS./HARDCOVER/$29.95/ISBN 3-540-63698-6

NANCY BAXTER HASTINGS, Dickinson College, Carlisle, PA

WORKSHOP CALCULUS WITH GRAPHING CALCULATORS

Guided Exploration With Review, Volume 1 and Volume 2

Developed as part of the well-known NSF-sponsored project, Workshop Mathematics, these texts are intended for use with students in a math laboratory, instead of a traditional lecture course. These new versions of the popular books, *Workshop Calculus,* are based on the use of graphing calculators by students enrolled in calculus. There is enough material to cover precalculus review, as well as first year single variable calculus topics. Student-oriented activities, experiments, and graphing calculator exercises are found throughout the text. Although these materials are graphing calculator dependent, there are no references in the text to a particular type of calculator. There is an appendix for use with a TI-83 which is available electronically and can be customized for use with other types of calculators.

Contents: Volume 1: Functions • Function Construction • Function Classes • Limits • Derivatives and Integrals: First Pass • Volume 2: Derivatives and Integrals: First Pass • Derivatives: The Calculus Approach • Definite Integrals: The Calculus Approach • Methods of Integration • Using Differentiation and Integration • Appendices: A. Using the TI-83 Graphing Calculator in Workshop Calculus • B. Integral Tables

VOLUME 1, with Sr. Barbara Reynolds: 1999/APP. 376 PP., 70 ILLUS./SOFTCOVER/$39.95 (TENT.)/ISBN 0-387-98636-7
VOLUME 2: 1999/APP. 424 PP., 65 ILLUS./SOFTCOVER/$39.95 (TENT.)/ISBN 0-387-98675-8

ALLEN HIBBARD, Central College, Pella, IA and **KENNETH LEVASSEUR,** University of Massachusetts, Lowell

EXPLORING ABSTRACT ALGEBRA WITH MATHEMATICA®

This upper-level laboratory supplement for courses in abstract algebra consists of several Mathematica packages, along with two collections of labs for group theory and ring theory. Additionally, there is a *Users Guide* which illustrates the functionality of the underlying code. Exercises occur in the stream of the text of the lab, which provides a context within which to answer. Questions are designed to be written either into the electronic notebook or on paper, whichever the instructor prefers. Flexibility of use, and the intention of the authors to make this work highly visual, e.g., with the inclusion of a full-color insert of significant algebraic concepts/images, make this publication pedagogically useful to both instructors and students alike. The multi-platform CD-ROM contains both Mathematica version 2.2 and 3.0 notebooks. There are also several Mathematica-produced animations aboard, along with MathReader, the public domain software which enables users to view content without having access to Mathematica itself.

TELOS 1999/496 PP., 20 ILLUS./SOFTCOVER BOOK, CD-ROM/$45.00/ISBN 0-387-98619-7

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036